Development of a new platform technology for the recognition and validation of peptide-protein interactions

Von der Fakultät für Lebenswissenschaften
der Technischen Universität Carolo-Wilhelmina
zu Braunschweig
zur Erlangung des Grades eines
Doktors der Naturwissenschaften
(Dr.rer.nat.)
genehmigte
D i s s e r t a t i o n

von Andrzej Swistowski
aus Klodzko / Polen

1. Referentin oder Referent: Prof. Dr. Jürgen Wehland
2. Referentin oder Referent: Prof. Dr. Stefan Dübel
eingereicht am: 13.02.2006
mündliche Prüfung (Disputation) am: 14.06. 2006

2006
(Druckjahr)

Vorveröffenlichungen der Dissertation

Teilergebnisse aus dieser Arbeit wurden mit Genehmigung des Fachbereichs für Biowissenschaften und Psychologie, vertreten durch den Mentor /Betreuer der Arbeit, in folgenden Beiträgen vorab veröffentlicht:

Tagungsbeiträge

Swistowski A, Bialek K, Beutling U, Frank R. Development of a new biochip platform technology for the recognition and validation of peptide-protein-interactions. 2nd BMBF-Kolloquium „New efficient methods for functional proteome analysis",
30. May - 1. June 2005, Potsdam (Poster & Abstract)

Beutling U, Bialek K, Gupte V, Swistowski A, Rübenhagen R, Frank R. A New Planar Support for Multiple Affinity Enrichment. DECHEMA Statusseminar Chip-Technologien, 3.-4. February, Frankfurt, (Poster & Abstract)

Bialek K, Swistowski A, Mühle L, Frank R. Development of a new biochip platform technology for the recognition and validation of peptide-protein-interactions. 3rd International and 28th European Peptide Symposium. Prague, Czech Republic, September 5-10, 2004. (poster and abstract).

Publikationen/Veröffentlichungen

Bialek K, Swistowski A, Beutling U, Mühle L, Frank R (2004) Development of a New Biochip Platform Technology for the Automated Recognition and Validation of Peptide-Protein-Interactions. In: Solid Phase Synthesis & Combinatorial Libraries 2004́, (Epton, R., ed.) Mayflower Worldwide Ltd., Kingswinford, UK, Chp. 10, 35-38.

Bialek K, Swistowski A, Frank R. (2003) Epitope-targeted proteome analysis: towards a large-scale automated protein-protein-interaction mapping utilizing synthetic peptide arrays. Anal Bioanal Chem., Aug, 376, 7:1006-13, Epub 2003 Apr 03.

Frank, R.; Bialek, K., and Swistowski, A. Large-scale protein-interaction mapping with synthetic peptide arrays: epitope-targeted proteome analysis. Peptides 2002; Proceedings 27th European Peptide Symposium, pp. 102-103.

Moim Rodzicom

Mojej żonie Ani

Table of Content

Abbreviations

Abbreviation	Name
A	Adenosine/Desoxyadenosine (-phosphate)
Ab	Antibody
Abl	Tyrosine-protein kinase ABL
Abp1	Actin binding protein 1
Acm	Acetamidomethyl protecting group
AFAP	Actin filament associated protein
Amp	Ampicillin
AP	Alkaline Phosphatase
APEG	amino PEG (Poly-ethylene-glycol) spacer
Ap^R	Ampicillin resistance
ARVCF	Armadillo repeat protein deleted in velo-cardio-facial syndrome
BCIP	5-bromo,4-chloro,3-indolylphosphate
BLAST	Basic Local Alignment Search Tool
Boc	t-Butyloxycarbonyl protecting group
bp	Basepair
BPB	Bromophenol Blue
BRCT	Conserved domain found at C-terminus of the breast cancer susceptibility gene
BSA	Bovine Serum Albumin
C	Cytidine/desoxycytidine (-phosphate)
C3G	Guanine nucleotide-releasing factor 2
c-ABL	Abelson murine leukemia viral oncogene homolog 1
CBS	Citrate-Buffered Saline
CD2	T-cell surface antigen CD2 precursor
CD2BP2	CD2 Cytoplasmic Domain binding protein
CDC 42 GAP	Cdc42 GTPase-activating protein
CDC42	Cell Division Control protein 42
cDNA	Complementary DNA
cfu	Colony Forming Unit
Cm	Chloramphenicol
Cm^R	Chloramphenicol resistant
COLT	Cloning of Ligand Targets
CORT	Cloning of Receptor Targets
Crk	Proto-oncogene C-crk (P38)
Cy3	Cyanine 3-dCTP
Cy5	Cyanine 5-dCTP
DCE	1,2-Dichloroethene
DCM	Dichloromethane
DIC	N,N'-Diisopropylcarbodiimide
DMF	N,N- dimethylformamide
DMSO	Dimethylsulfoxide
DNA	DeoxyriboNucleic Acid

Abbreviation	Name
dNTP	Deoxyribonucleoside Triphosphate
ds	Double stranded
DTT	Dithiothreitol
Dvl1	Segment polarity protein dishevelled homolog DVL-1
E. coli	*Escherichia coli*
EDTA	Ethylendiamintetraacetate
EH	Eps15 homology domain
ELISA	Enzyme Linked ImmunoSorbent Assay
ErbB-4	Tyrosine kinase erbB-4
EVH1 (E)	Ena-VASP Homology domain 1
Ezr	Ezrin
f.c.	Final concentration
F_2Pmp	(phosphonodifluoromethyl)-phenylalanine
FACS	Fluorescence Activated Cell Sorting
Fas	Apoptosis-mediating surface antigen FAS
FE65 (F)	Amyloid beta A4 precursor protein-binding family B member 1
FF	A domain with two conserved phenylalanines (F);
FHA	Forkhead-associated;
Fmoc	Fluorenylmethyloxycarbonyl
Fyn	Proto-oncogene tyrosine-protein kinase Fyn
G	Guanosine/Desoxyguanosine (-phosphate)
GBF	Gesellschaft für Biotechnologische Forschung
GST	Glutathione S-transferase
GYF	Glycine-tyrosine-phenylalanine protein domain
h	Hour
HAc	Acetic Acid
HeLa cells	Cell line taken from Henrietta Lacks alias Helen Lane
hnRNP K	Heterogeneous nuclear ribonucleoprotein F
HPLC	High Performance Liquid Chromatography
Hrb	HIV-1 Rev-binding protein
HRP	Horse Radish Peroxidase
HUGE	A Database of Human Unidentified Gene-Encoded Large Proteins analyzed by Kazusa human cDNA project
IL-4	Interleukin-4
IPTG	Isopropyl-β-D-thiogalactopyranoside
kb	Kilobase
KD	Dissociation constant
KDa	Kilodalton
KH	Ribonucleoprotein K homology
Km^R	Kanamycin resistance
L. monocytogenes	*Listeria monocytogenes*
L.M.P agarose	Low Melting Point agarose
LB	Luria-Bertina medium

Abbreviation	Name
LB-EBV	B Lymphocyte cells immortalized by the Epstein-Barr virus;
Lin1p	Protein LIN1
MALDI-MS	Matrix Assisted Laser Desorption Ionization Mass Spectrometry
Mena	Enabled protein homolog
min	Minute
MTT	Methylthiazolyldiphenyl-tetrazolium-bromide
NMP	1-Methyl-2-pyrrolidinone
NUMB	Protein numb homolog
OD	Optical density
OPAL	Oriented Peptide Array Library
OPD	o-Phenylene diamine
OtBu	t-butyl ester
p.a.	Per analysis
$p130^{Cas}$	CRK-associated substrate
PBS	Phosphate Buffer Saline
PCR	Polymerase Chain Reaction
PDZ	Protein domain also called Discs-large homologous regions (DHR)
PEG	Polyethylene Glycol 6000
pfu	Plaque Forming Unit
PH	Pleckstrin homology
PI3K p85	PI3-kinase p85-alpha subunit
Pmc	2,2,4,6,7-pentamethyldihydrobenzofuran-5-sulfonyl protecting group
PMSF	phenylmethanesulfonyl-fluoride
POD	Peroxidase
PPII	Polyproline Type II helix
Prp-8	Pre-mRNA splicing factor PRP8 homolog
PTB	Phosphotyrosine-binding
rpm	Rounds per minute
RRT	Recombinant RT-loop
RSV	Rous sarcoma virus
RZPD	German Resource Center for Genome Research
SDS	Sodium Dodecyl Sulfate
sec	Second
SETA	Src homology 3 expressed in tumorigenic astrocytes;
SF 1	Steroidogenic Factor 1
SH 1, 2, 3	Src homology 1, 2, 3
Sh3yl1	Sh3 domain YSC-like 1
Shb	Ubiquitously expressed Src homology 2 protein
SmB	snRNP-associated protein
SOS-1	Son of sevenless protein homolog 1
SPR	Surface Plasmon Resonance
SSC	Sodium chloride-Sodium Citrate buffer
STRAP	Streptavidin Alkaline Phosphatase complex

Abbreviation	Name
Str^R	Streptomicin resistant
Swiss-Prot	Protein knowledgebase
T	Thymidine
TAE	Tris/Sodium acetate/EDTA
TAIS	Target-Assisted Iterative Screening
TBE	Tris-Borate-EDTA
TBS	Tris Buffered Saline
TBST	Tris Buffered Saline -Tween
td	Theoretical degeneracy
TE	Tris/EDTA
Tet^R	Tetracycline resistance
TFA	Trifluoroacetic acid
TIBS	Triisobutylsilane
TrEMBL	Computer-annotated supplement to Swiss-Prot
Tris	Tris-(hydroxymethyl)-aminomethane
Trt	trityl protecting group
U	Unit
UD	The N-terminal unique domain of Src kinases;
UV	Ultra Violet
V	Volt
WBP	WW domain-binding proteins
WISE	Whole Interactome Scanning Experiment
WW	A domain with two conserved tryptophans (W)
WWP1	WW domain-containing Protein 1
YAP (Y)	65 kDa Yes-associated protein

DNA codes

G	Guanosine	S	C or G, Strong (3 H-bonds)
A	Adenosine	M	A or C, aMino
T	Thymidine	B	C, G, or T (i.e., not A)
C	Cytidine	D	A, G or T (i.e., not C)
R	G or A, puRine	H	A, C or T (i.e., not G)
Y	Cor T, pYrimidine	V	A, C,or G (i.e., not T)
K	G or T, Keto	N	A, C, G or T, aNy
W	A or T, Weak (2 H-bonds)	X	A, C, G,or T

Amino Acid Abbreviations

Amino Acid	Three Letter	One Letter
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Asparagine/aspartic acid	Asx	B
Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamic Acid	Glu	E
Glutamine	Gln	Q
Glutamine/glutamic acid	Glx	Z
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

Symbols representing consensus sequences of peptide ligands and protein domains

According to Aasland et al., 2002

Brief summary of rules	Symbol	Brief summary of rules	Symbol
Unknown, other or any amino acid	x	Phosphorylated amino acid (Tyr)	*po*Y
Sequences flanking ligand core at the N-terminus	*fn*	Sulfated amino acid (Tyr)	*su*Y
Sequences flanking ligand core at the N-terminus	*fc*	O-glycosylated amino acid (Ser)	*gl*S
Hydrophobic amino acid	ϕ	N-glycosylated amino acid (Asp)	*gl*N
Aromatic amino acid	Ω	Methylated amino acid (Arg)	*me*R
Hydrophilic amino acid	ξ	Symmetrically methylated Arg	*sme*R
Positively charged residue	[+]	Asymmetrically methylated Arg	*ame*R
Positively charged residue	[-]	Acetylated amino acid (Lys)	*ac*K
Aliphatic side chains	ψ	Hydroxylated Proline	*hy*P
Small chain amino acid	π	Excluded amino acid (Pro)	^P

1. Introduction

1.1 Peptide and protein repertoires for global analysis of modules

Protein-protein interactions are responsible for most of the cellular processes and control the cell physiology. Cellular processes are thought to be regulated rather by complicated protein complexes and protein networks than by single proteins (Fig 1.1.1). Disruption of some sensitive connection within interaction network may drive the cell into a pathological state. Despite availability of the sequences of entire genomes, the function of many proteins remains unknown. Studies concerned with the relationship between proteins within complexes may help to address their biological functions, which is a primary goal of functional genomics and proteomics. This is supported by high-throughput technologies allowing for global analysis of the entire proteomes (Uetz *et al.,* 2000; Ho *et al.,* 2002; Gavin *et al.,* 2004).

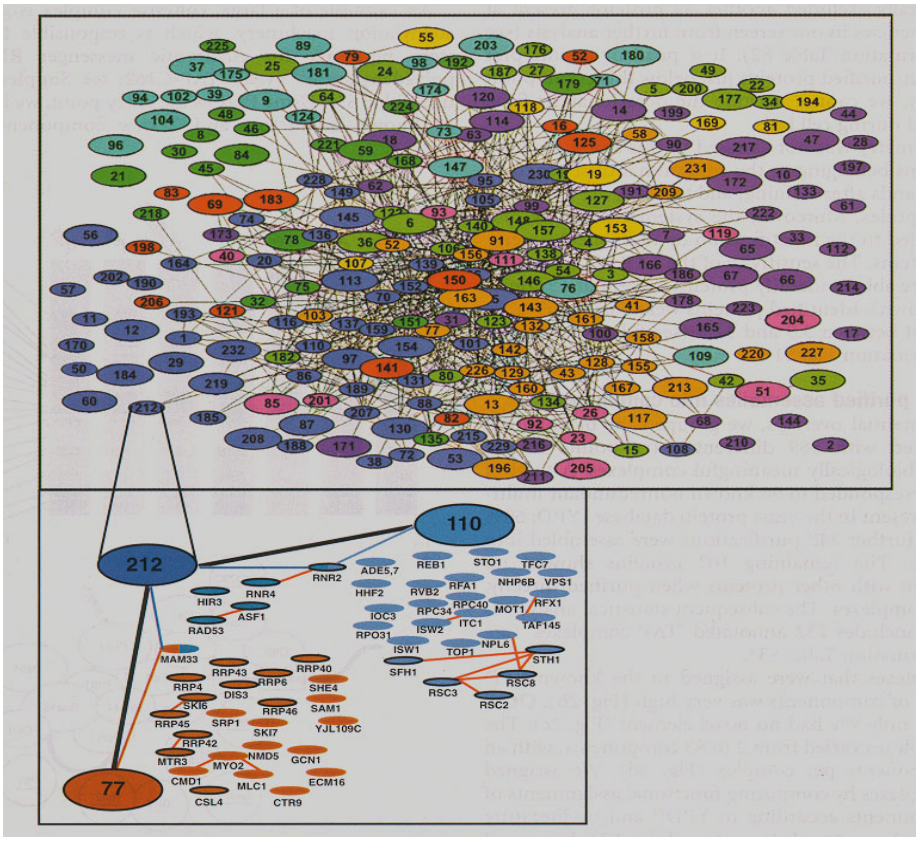


Fig 1.1.1: The protein complex network (Gavin *et al.,* 2004)

Many proteins, most prominently those of regulatory function, are built from smaller domains, which are stably folded structural modules, still displaying their specific functional property. A substantial fraction of protein interactions is mediated by these modules, recognizing and binding to short peptides in the partner protein. The catalogue of protein domains is rapidly growing (kringel, SH2, SH3, PH, EVH1, PDZ, WW, etc. (Cesareni *et al.,* 2004), indicating a more general principle utilized by nature. Protein domains are found in proteins involved in the regulation of cellular events such as signal transduction, the cell cycle, protein trafficking, targeted proteolysis, cytoskeletal organization and gene expression. Therefore understanding of module-ligand recognition mechanisms, relationship between interacting partners as well as identification and classification of new protein modules is of high importance. In this chapter an overview is given on the experimental approaches that have been conceived and are being used to search genomes and proteomes for functional protein domains and their ligands as well as to assess their recognition specificities and functions on a molecular basis. With respect to the empirical nature of search approaches, most of these are very much concerned with the intelligent design and sophisticated preparation of competent domain and ligand repertoires that are suited to represent natural global pools of potential structures and which allow for efficient experimental exploration. A particular attention was paid on the experimental design and concepts of the reports considered for this chapter, as well as literature reference for more information on the types of domains and ligands that have been discovered this way. This section is organized according to the origin of the domain/ligand repertoires reported and the principles applied to prepare and screen them. This involves cell biology, biochemistry, molecular biology, and chemistry, which illustrates the interdisciplinary nature of the work in this field of research. The recent efforts in developing non-natural and small molecule drug-like ligands for use in therapeutic interference with protein domain function are not covered in this chapter.

1.1.1 Repertoires from cell extracts

A cell extract is the most natural source for a repertoire of the proteins to be investigated for function as modules or ligands. Cell extracts are relatively easy to prepare and their advantageous feature is that all proteins are correctly folded and post-translationally modified. Moreover, proteins in cell extracts may be easily metabolically labeled, for example, with ^{35}S or ^{32}P. Usually, protein extracts are probed with individual module or ligand molecules, and binding partners are isolated by affinity chromatography or pull-down of complexes via a tag at the probe molecule (Weng *et al.,* 1993; Weng *et al.,* 1994; Ikeda *et al.,* 2000) or by immunoprecipitation (Liu *et al.,* 1999). The separated proteins are fractionated on polyacrylamide gels (one-or two-dimensional electrophoresis) and identified

by mass spectrometric peptide fingerprinting, or they may be blotted and inspected with, e.g., antibodies. Such approaches were recently applied in a high-throughput systematic way to analyze the entire yeast proteome and yielded impressively comprehensive protein interaction networks (Ho *et al.,* 2002; Gavin *et al.,* 2004). These networks obviously include many protein module interactions and may reveal novel module/ligand families.

1.1.2 Repertoires of proteins based on expression cloning of DNA libraries

The major source for global domain and ligand repertoires is heterologous expression of coding DNA libraries. A common theme of approaches based on such sources is to construct a physical link between the individual member of the protein repertoire and its coding DNA fragment, as well as a suitable selection process for the module-ligand complexes. Usually, the complex partners are called bait and prey; either one or both may be represented by a global library. A “DNA tag” always allows for efficient, easy amplification of even extremely minute amounts of module/ligand molecules. Identification is achieved through DNA sequence analysis (DNA sequencing or hybridization with respective complementary DNA probes). Examples are the yeast two-hybrid system (Allen *et al.,* 1995) and the split-ubiquitin analog (Johnsson and Varshavsky, 1994) phage expression library screens (Sambrook *et al.,* 1989), the various molecular display systems (Allen *et al.,* 1995; Fish, 2001) used in iterative rounds of affinity enrichment (also called biopanning), and more recently, large-scale expressed protein collections in microtiter plate or microarray formats (Mac Beath, 2002). cDNA libraries are the most suitable source of global protein repertoires, rarely available in full length, but complementing the full proteome in the form of shorter redundant libraries of fragments. cDNA libraries can be chosen to represent specific cellular states, such as fetal or adult, brain or liver, healthy or oncogenic/pathogenic, etc. Genomic DNA libraries have been used only with simple model organisms such as yeast, *Caenorhabditis elegans*, etc. but are not considered for higher organisms, which carry increasing amounts of non-coding genetic material. A limitation of heterologous expression is the lack of proper posttranslational modification, which in some instances can be overcome by using genetically engineered host strains and in others may be introduced in vitro. Recently, global studies were reported that investigated the protein interactomes of whole organisms nonbiased for domain–ligand interactions. These have established systematic proteome interaction maps utilizing large-scale yeast two hybrid screens for yeast itself, nematode, and *Drosophila* (Ito *et al.,* 2001; Giot, 2003). Work on the human interactome is underway. Together with the global protein complex analyses mentioned above (Section 1.1.1), comprehensive catalogs of protein interactions are currently being assembled. These will be mined with bioinformatics tools for known and new module-ligand interactions.

1.1.2.1 Ligand repertoires used with the yeast two-hybrid system

The DNA for a module of interest is cloned into a bait expression plasmid in-frame with the gene for the DNA binding domain of a transcription factor, and a library of cDNA or fragmented genomic DNA is cloned into a prey expression plasmid in frame with the gene for the transcription activation domain of that transcription factor. Yeast transformants resulting from transfection or mating which carry both plasmids and produce active interacting partners are identified by activation of a reporter gene such as *lacZ* or a nutritional marker under control of the reconstituted transcription factor. Fig 1.1.2 shows the schematic view of this process.

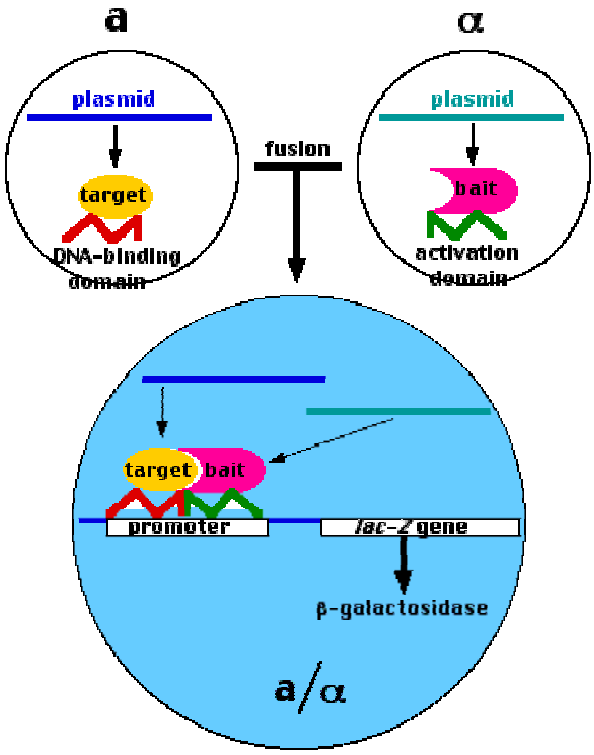


Fig 1.1.2: Schematic representation of the principle of yeast two-hybrid system. Picture source: http://users.rcn.com/jkimball.ma.ultranet/BiologyPages/T/Two_Hybrid.gif

Several examples are listed in (Table 1.1.1).Tong *et al,* exploited yeast two-hybrid screens with prey libraries in various formats: a genome-wide array of transformants representing all yeast open-reading frame fusions and conventional yeast genomic and cDNA libraries (Tong *et al.,* 2002). The screens were used to systematically confirm module–ligand interactions, which had been derived from consensus binding motifs and computational network analysis based on phage-display screens of random peptide libraries (see sections: 1.1.4.1 and 1.1.4.2). This approach was applied to yeast SH3 domains and used to verify 59 interactions

for 39 proteins by matching a phage-display network of 394 interactions among 206 proteins and a yeast two-hybrid network of 233 interactions among 145 proteins.

Table 1.1.1: Screens of ligand repertoires with the yeast two-hybrid system.

Prey library	Bait	Reference
HeLa cDNA	SH2SH3 of c-ABL	Zhu and Shore, 1996
human fetal brain cDNA	all 3 SH3 of Nck	Matuoka *et al.,* 1997
human brain cDNA	SH3 of αI spectrin	Ziemnicka-Kotula *et al.,* 1998
human placenta cDNA	SH3 of $p130^{Cas}$	Kirsh *et al.,* 1999
human mammary gland cDNA	bovine SF1	Zhu *et al.,* 2000
p53 - astrocyte cDNA	SH3-1 and SH3-23 of SETA	Borinstein *et al.,* 2000
yeast genomic DNA	SH3 of Abp1	Fazi *et al.,* 2002
mouse skin cDNA	SH3 of Sh3yl1	Shimomura *et al.,* 2003
murine keratinocyte cDNA	Fyn	Seykora *et al.,* 2002
human lung fibroblast cDNA	UD/SH2/SH3 of c-Src	Chang *et al.,* 1998
T-cell cDNA and 11.5-day mouse embryo	human p52 Shc	Schmandt *et al.,* 1999
human brain cDNA	WW1 of YAP	Espanel *et al.,* 2001
human leukocyte cDNA	EVH1of Mena(S)	Tani *et al.,* 2003

1.1.2.2 Module repertoires used with the yeast two-hybrid system

In the reverse approach, when cloning a ligand of interest into a bait plasmid, the same prey libraries can be used to select for modules. Several examples are listed in Table 1.1.2.

Table 1.1.2: Screens of module repertoires with the yeast two-hybrid system.

Prey library	Bait	Selected domain	Reference
LB-EBV cDNA	$p47^{phox}$	SH3	Fuchs *et al.,* 1995
mouse brain cDNA	Alix	SH3	Chatellard-Causse *et al.,* 2002
rat brain cDNA	proline-rich tail (aa 1015-1308) of the 145-kDa isoform of synaptojanin	SH3	Ringstad *et al.,* 1997
rat brain cDNA	N-terminal part (1-65 aa) of Pγ-rod	SH3	Morin *et al.,* 2003
human brain cDNA	cytoplasmic region (aa 790-1356) of KDR	SH2	Igarashi *et al.,* 1998
human fetal brain cDNA	P413 to P713 fragment of athrophin-1	WW	Wood *et al.,* 1998
Activated human T cell cDNA	cytoplasmic region (aa 221-327) of CD2	GYF	Nishizawa *et al.,* 1998

Prey library	Bait	Selected domain	Reference
lung cDNA library	51 aa of the C-terminus of rat CFTR	PDZ	Cheng *et al.,* 1998

1.1.2.3 Phage expression libraries of ligands

The genomic or cDNA library is cloned primarily into a lambda expression vector, packaged into phage particles, and plated onto a bacterial lawn. The phage plaques producing the fusion protein are lifted onto nitrocellulose filters, and these are probed with the labeled or tagged protein module of interest. DNA of positive individual phage plaques is then sequenced. Table 1.1.3 gives an overview of reported screens.

Table 1.1.3: Screens of phage expression libraries of ligands

Vector	Prey library	Bait	Reference
λEXlox	16-day mouse embryo cDNA	GST-SH3 of c-Abl, cCrk, c-Src, n-Src	Alexandropoulos *et al.,* 1995
λEXlox	16-day mouse embryo cDNA	GST-Gads	Berry *et al.,* 2002
λEXlox	16-day mouse embryo cDNA	^{35}S-labled GST-SH3-SH3 of Crk	Schumacher *et al.,* 1995
λEXlox	16-day mouse embryo cDNA	^{32}P-labelled GST-WW of YAP	Chen *et al.,* 1995
λEXlox	16-day mouse embryo cDNA	EH-AP of Intersectin	Yamabhai *et al.,* 1998
λEXlox	12-, 14-, 16-day mouse embryo cDNA	^{32}P-labelled GST-WW of FE65	Ermekova *et al.,* 1997
λgt11	mouse pre-B cell line 22D6 cDNA	biotinylated GST-SH3 of Abl	Cicchetti *et al.,* 1992
λgt11	human spleen cDNA	GST-SH3 of CRK	Tanaka *et al.,* 1994
λZAP	macrophage cDNA	GST-SH3 of Nck	Rivero-Lezcano *et al.,* 1994

1.1.2.4 Phage expression libraries of modules

The lambda phage-based cDNA libraries mentioned above have also been used to search for proteins containing modules, by probing the filter lifts with labeled ligands such as synthetic peptides or fragments of proteins that carry functional sites. Several examples are listed in Table 1.1.4.

Table 1.1.4: Screens of phages expression libraries of modules

Vector	Pray library	Bait	Selected domain	Reference
λEXlox	16-day mouse embryo cDNA	biotin-SGSGSRLTQSKPPLPPKPSWVSR-NH_2 Cortactin SH3 ligand from phage library (Aparks et al., 1996)	SH3 [a)]	Sparks *et al.,* 1996
λEXlox	10.5-day mouse limb bud embryo cDNA	$(His)_6$-proline-rich fragment of formin (518-750 aa)	SH3, WW	Chan *et al.,* 1996
λEXlox	16-day mouse embryo cDNA	mixture of biotinylated DHQ*po*Y*po*YNDFPGKEPP, DHQ*po*YYNDFPGKEPP DHQY*po*YNDFPGKEPP DHQYYNDFPGKEPP ELFDDS*po*YVNIQNLD derived from human p52 Shc	SH2 [b)]	Liu *et al.,* 1998
λTriplEX	human brain cDNA	mixture of biotinylated PGTP_4YTVGPGY (WBP-1), YVQP_4YPGPM (WBP-2A), PGTPYP_4EFY (WBP-2B), G_3FPPLP_4YPPLG (Ras-GAP)	WW [a)]	Pirozzi *et al.,* 1998
λgt11	chicken embryo cDNA	proline rich peptide comprised of three tandem copies of the 11-aa from the RSV Gag p2b sequence	WW	Kikonyogo *et al.,* 2001

[a)] this approach was termed “cloning of ligand targets” (COLT)

[b)] this approach was termed “cloning of receptor targets” (CORT)

1.1.2.5 Bio-Panning with phage displayed protein ligands and domains

Technology

Since its invention (Smith, 1985) affinity enrichment of peptide/protein binders using phage display technology was successfully applied in the immunology, cell biology, drug discovery and pharmacology research fields and still remains very attractive, especially when combined with high-throughput approaches. Physical connection of the peptide/protein displayed by fusing them to a phage capsid protein (phenotype) with the nucleic acid genome of the phage particle (genotype) is the principle feature of phage display allowing for an easy selection and identification of target specific binders. This is achieved by genetic fusion of the nucleotide sequences encoding the proteins/peptides to be displayed, in frame to the gene encoding a phage coat protein. Huge repertoires, up to 10^{11} candidates, can be

generated by cloning of e.g. gene fragments, random oligonucleotides and cDNA populations into phage chromosome. Typically, ligand specific binders are selected in an affinity process called “bio-panning”. This process is usually repeated several times and results in enrichment of partners that strongly interact with their ligand. For description see Fig 1.1.3. Several different phage-display systems utilizing filamentous (M13) and virulent (T7, T4, lambda) phages were developed. A very informative comparison of their particular properties was given by Castagnoli (Castagnoli *et al.,* 2001).

Phage display of protein ligands

DNA libraries are cloned into phage vectors in-frame with one of the phage coat protein genes. The resulting phages display the fusion protein on their surface and are used to select binding ligands by affinity enrichment on immobilized modules. Several interesting reports utilizing this approach are described below; however, it is more extensively used with repertoires of small peptide ligands (see section 1.1.4). Ligands for SH3 domains were isolated from phage-displayed expression libraries of cDNA from the human epithelial breast carcinoma cell line MCF7 and a murine fibroblast NIH 3T3 cDNA using the T7Select1-1b system. Both libraries (10^7 primary recombinants with an average length 0.7–0.8 kb) were screened with the human Grb2-GST (GST = glutathione S-transferase) fusion protein as bait (Zozulya *et al.,* 1999). A fragmented yeast genomic DNA library expressed and displayed as N-terminal fusion to the pVIII protein of M13 phage (10^8 independent clones representing more than 50 times the whole yeast genome and presenting yeast protein fragments of average length 40 amino acids) was affinity-selected for ligands bound by the full length yeast Abp1 protein, which was expressed as GST fusion and immobilized to glutathione–*S*-Sepharose (Fazi *et al.,* 2001). A fragmented human leukocyte cDNA library (10^8 clones, 0.2–0.5 kb) was fused to the bacteriophage M13 pIII gene. This library was first depleted of non-phosphorylated ligands by passage over Sepharose loaded with the SH2 domains of SHP-2, then was phosphorylated in vitro with fyn kinase and panned against the tandem SH2 domains of SHP-2 (Cochrane *et al.,* 2000). Usually, several rounds of panning with intermediate amplification of phage eluates by bacterial infection are performed to enrich over unspecific binders. Kurakin combined phage display and expression library screening formats to overcome this drawback (Kurakin and Bredesen, 2002). After the first panning round, eluted phages are plated onto a bacterial lawn, and plaques are transferred onto a nitrocellulose membrane. The protein target tagged with a reporter such as alkaline phosphatase is then used as a one-step detection reagent to screen for interacting plaques on the membrane. They named this procedure TAIS (target-assisted iterative screening). Using this method with a normal human brain T7 phage-displayed cDNA library and several

bait modules (PDZ domain of PSD95; SH3 domains of Src, Abl, and Crk; third WW domain of Nedd4), they identified 12 novel putative and two previously.

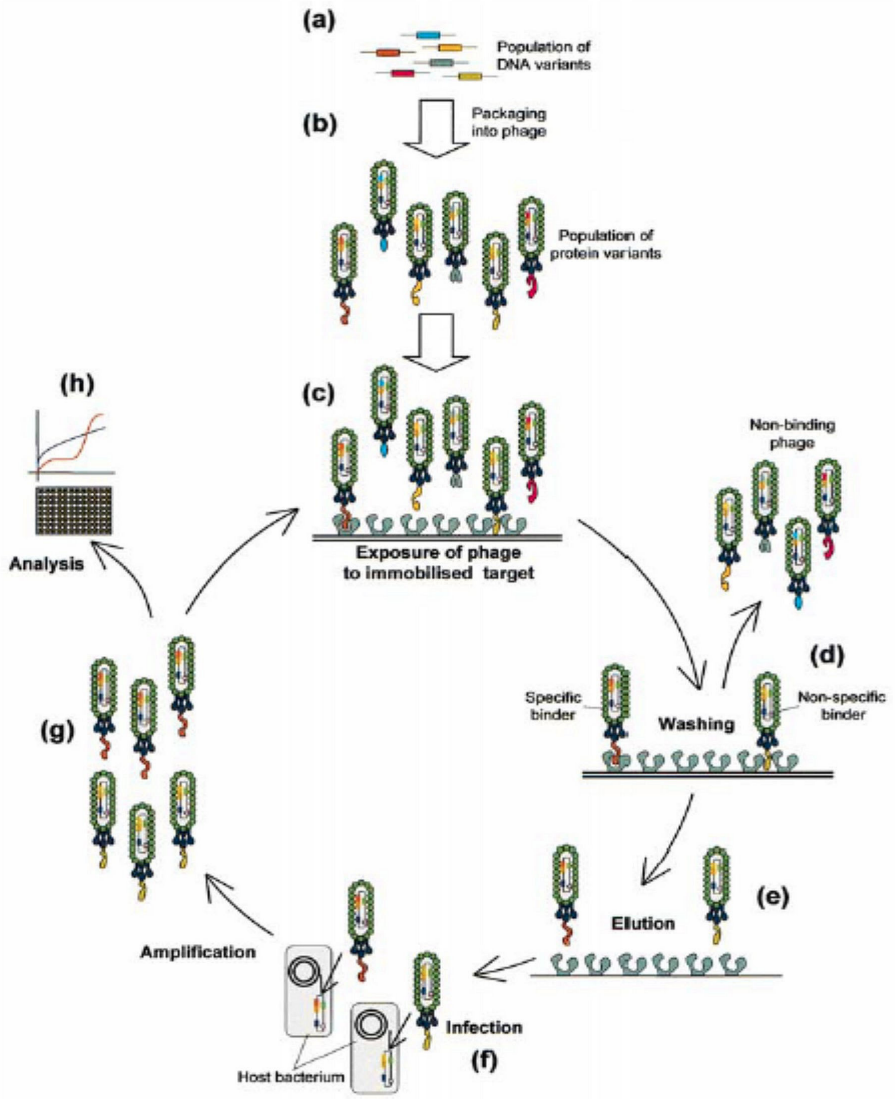


Fig 1.1.3: The phage display cycle (a) A library of variant DNA sequences encoding proteins or peptides is created and (b) cloned into phage or phagemid genomes as fusions to a coat protein gene. (c) The phage library displaying variant proteins or peptides is exposed to target molecules and phages with appropriate specificity are captured. (d) Non binding phages are washed off – although some non-specific binding may also occur. (e) Bound phages are eluted by conditions that disrupt the interaction between the displayed peptide or protein and the target. (f) Eluted phage are infected into host bacterial cells and thereby amplified. (g) This amplified phage population is in effect a secondary library that is greatly enriched in phage displaying peptides or proteins that bind to the target. If the bio-panning steps (c) to (f) are repeated the phage populated becomes less and less diverse as the population becomes more and more enriched in the limited number of variants with binding capacity. (h) After several (usually three to five) rounds of bio-panning monoclonal phage populations may be selected and analyzed individually, (Willats, 2002).

Phage display of protein domains

Again, in the reverse of the scheme above, protein modules have been searched for by panning phage displayed cDNA libraries on immobilized ligand baits. SH3 and WW domains were isolated from an expression cDNA library made from human brain displayed on the lambda D capsid protein (9.8 x 10^7 primary recombinants) by panning on a GST fusion protein of the proline-rich fragment (residues 1058 – 1119) of synaptojanin 1 (Zuconni *et al.,* 2001). WW domains were isolated from a normal human brain T7 phage-displayed cDNA library screened with the WW-YAP ligand peptide GTP4YTVG, which was tethered to a spot on a cellulose membrane. This method will be further advanced into an automated process utilizing SPOT arrays of synthetic peptide libraries as multiple baits to systematically scan fragments from hundreds of proteins for interacting modules (Bialek *et al.,* 2003). The phosphotyrosine-binding domains SH2 and PTB were isolated from MCF7 and NIH 3T3 phage-displayed libraries (for details see section 1.1.2.5), which were screened with the EGFR- and Shc-derived synthetic phosphopeptides (EGFR-pY1068 biotin-DDAFLPVPEpoYVNQSVPKR, EGFR-pY1148 biotin–SHQMSLDNPD*po*YQQDFFPK, and Shc-pY317 biotin– LFDDPS*po*YVNVQNL) as baits (Zozulya *et al.,* 1999).

1.1.2.6 Protein arrays as ligand repertoires

The availability of fully sequenced and annotated genomes of model organisms open up the opportunity to prepare comprehensive sets of individually expressed and purified proteins as bait or prey repertoires. This has been achieved for 5800 open reading frames of *Saccharomyces cerivisia* and from these a yeast proteome microarray containing approximately 80 % of the yeast proteins was constructed. The yeast proteins were fused to GST-Hisx6 tags at their N-termini and spotted onto nickel-coated glass slides. These arrays were probed e.g. with calmodulin in the presence of calcium. Six of the twelve known calmodulin targets were detected. Of the remaining six, two were not present in this collection and four were not detectable with anti-GST antibodies. 33 additional potential partners for calmodulin were identified. Sequence comparison revealed that 14 of the 39 calmodulin-binding proteins contain a motif whose consensus is *fn*(I/L)QxxK(K/x)G[+]*fc* (Zhu *et al.,* 2001).

1.1.2.7 Protein arrays as domain repertoires

A repertoire of 212 GST fusion proteins and protein domains was collected from different sources: 33 WW, 23 SH3, 17 SH2, 23 PH, 23 PDZ, 8 FF, 7 14.3.3, 5 PTB, 4 FHA, 2 KH and 67 proteins without canonical domain structure. These were arrayed onto nitrocellulose-

coated glass slides to produce a protein-domain chip. Five peptides, including one with methylated arginine residues as an example for a post-translational modification, all with known binding specificity to the domains above, were used to confirm that the arrayed proteins retain their binding integrity. Extract of human HCF7 cells and antibodies against Sam 68 and SmB´, as well as peptides derived from these proteins were used to demonstrate distinct and reproducible binding to subsets of this array. In this manner, module-specific profiles for cellular proteins were obtained from the binding patterns. These were at most similar to those obtained with the respective peptide ligands and only slightly broader, which is to be expected since proteins may harbour additional interacting residues (Espejo *et al.,* 2002). 65 human WW domains were expressed as GST fusion proteins and immobilized onto the surface of the wells of a microtiter plate (Hu *et al.,* 2004). A quantitative ELISA-like binding assay was applied to screen these modules sequentially with 1930 putative synthetic peptide ligands that were selected by searching the Swiss-Prot and TrEMBL protein databases. A cross-affinity matrix that represented all the possible combinations of interactions between the human WW domains and their peptide ligands was established which revealed clear ligand preferences for the WW domains: 31 of them were classified to a subgroup that recognize the *fn*PPxY*fc* motif (elm:LIG_WW_1), 11 of them recognize *fn*PPLP*fc*/*fn*PPR*fc* (elm:LIG_WW_2/; elm:LIG_WW_3), three bind *fn*PGM*fc*/*fn*PPR*fc* (elm:LIG_WW_3) and four could not be classified to any subgroup.

1.1.3 Mutagenized domain libraries

Site directed mutagenesis with mutagenic oligonucleotide primers or random mutagenesis by error-prone PCR were applied to generate repertoires of domain variants. These repertoires were used to investigate the contribution of individual residues to domain activity and/or to search for domains with new properties. Obviously, all above listed methods for DNA library screens are applicable.

1.1.3.1 Site-directed mutagenesis

Two bacterial expression libraries of mutated YAP WW1 domains with random sequences at the position corresponding either to L190 and H192 together (td = 400) or Q195 alone (td = 20) were constructed in the pGEX-2TK. Colony lifts on nitrocellulose filters were screened with radiolabeled ligands of YAP (PPXY) or FE65 (PPLP). A single mutation L190W of the YAP WW1 domain was found to allow binding to both types of ligands and an additional substitution H192G almost completely shifted the selectivity to that of FE65 (Espanel and Sudol, 1999). A library of Hck SH3 domains (1.3×10^8 recombinant clones)

with a random hexapeptide replacement in the RT-loop (termed RRT-SH3) was generated by PCR-assisted mutagenesis using a degenerate primer. This library was displayed on the surface of M13 bacteriophage and used for panning with HIV-1 Nef and Nef-F90R mutant. RRT-SH3 domains that can bind to Nef up to 40-fold more avidly than Hck-SH3 were identified. Also RRT-SH3 domains with an opposite specificity for the Arg90 Nef-mutant were selected (Hiipakka *et al.,* 1999). A repertoire of 10^7 different SH3 domains was created by grafting the residues that are represented at the binding surfaces of natural SH3 domains onto the scaffold of the human Abl-SH3 domain (td ~ 2 x 10^9). The library was displayed by fusion to the C-terminus of the D protein of bacteriophage λ and screened by affinity selection with APTYP_3LPP and LSSRPLPTLPSP peptides, which are known ligands for human Abl and Src SH3 domains respectively. As few as two or three amino acid substitutions can lead to dramatic changes in recognition specificity (Panni *et al.,* 2002).

1.1.3.2 Random mutagenesis

A yeast reverse two-hybrid system (interactions activate a toxic reporter) was used to select single amino acid changes that abolish protein-protein interaction. A prey library of the mutated fragment (159-437) of the E2F1 transcription factor was prepared by a mutagenic PCR and screened with the full-length DP1 protein as bait in a two-step selection process which deselected misfolded variants in the second step. A putative helix region that is conserved among E2F family members was identified as the essential part of the E2F/DP1 heterodimerization domain (Vidal *et al.,* 1996). The PDZ domain of mouse AF-6 was mutated using error-prone PCR. This library was screened using the Y2H-system with four different C-terminal expressed peptides (-FHAALGAV, -FHAALGKYV, -FHTQITRYV and -SPANIYYKV) as baits. All these peptides do not interact with the native AF-6 PDZ domain. The majority of mutations leading to high affinity binding to the new peptide ligands were localized in the three secondary structure elements of the PDZ domain that are known to make up the peptide binding groove (Schneider, 1999). Four Pex5 protein single point mutants that abrogate binding to the SH3 domain of Pex13 were used as baits in Y2H-screen for suppressor mutants from a randomly mutagenized (error-prone PCR) prey library of the Pex13p SH3 domain. Mutations that restore the interaction with the Pex5, were found at a new unique site on the Pex13-SH3 domain that is distinct from the classical PXXP-binding pocket (Barnett *et al.,* 2000).

1.1.4 Repertoires of peptide ligands based on expression cloning of oligonucleotide libraries

A typical feature of the protein modules discussed in this section is their action either by binding or post-translational modification (including cleavage) on short target sequences (functional sites) present in other proteins. Such functional sites can be mimicked by respective short peptide sequences of completely synthetic origin. This section recapitulates approaches to generate such synthetic repertoires by cloning the corresponding coding double-stranded DNA repertoires made from synthetic oligonucleotides. The common codon scheme NNK or NNS is used for the chemical synthesis of degenerate oligonucleotide sequences where N is an equimolar mixture of A+C+G+T, K of G+T and S of G+C. In this way, the codons encoding all 20 amino acids are included and stop codons are restricted to TAG only (Adey *et al.,* 1996). A triplet (codon-based) chemical synthesis of random oligonucleotide repertoires is feasible and could avoid the expression of many nonsense peptides (Sondek *et al.,* 1992), but is only rarely used because the triplet building blocks and their chemical assembly are not trivial. Often codons for cysteines are included at predefined positions to allow for a cyclization of the expressed peptide structures to form constrained loops. These oligonucleotide repertoires are primarily expressed as fusions to larger proteins, with the phage display clearly dominating (see also section 1.1.2.5). The majority of libraries are based on the filamentous phage M13 with the fusion to the N-termini of either the minor coat protein pIII (5 copies at the tip of the phage particle) or the major coat protein pVIII (approximately 2700 copies forming the body of the phage particle). These two variations are particularly distinct in their exploitation or prevention of avidity effects in the selection of weak or high affinity binders. Other types of phages are rarely used for peptide display except when constructing fusions to the C-termini of proteins which is a prerequisite, for example, when selecting binders to PDZ domains. The D-protein of the lambda phage capsid and the 10B coat protein of the T7 phage are two examples. The T7 display system was also reported in the context of a novel target assisted iterative screening method (TAIS) as described already in section 1.1.2.5. As a result, a novel unconventional SH3 binding motif, *fn*Px(P/A)xxR*fc* was discovered and further characterized (Kurakin *et al.,* 2003). The proper display of peptides at the C-terminus of M13 phage coat protein VIII could be also achieved with the help of a special thirteen amino acid long peptide linker. Fusion to the C-terminus of the Escherichia coli Lac repressor was exploited for a plasmid display system by Sticker and collogues (Stricker *et al.,* 1997). Following bacterial expression, the Lac repressor protein binds to the *lac* operator sequence on the same plasmid, thereby linking each peptide to the plasmid encoding that peptide. Because of the small particle size, much higher diversities of peptides can be screened with such libraries.

1.1.4.1 Random peptide libraries

Random peptide libraries are of a generic nature and applicable to any screening task. Therefore, many random libraries of different structure concerning the length of the peptide insert and constraints by cysteine bridges that were constructed for other studies were also utilized for selecting ligands for protein modules. One example is the M13 pVIII displayed nonapeptide library X_9 by Felici (Felici *et al.,* 1991), which was widely used by many researchers for protein domain-peptide ligand interaction studies. Table 1.1.5 gives an overview of several reports found in the literature. The library of Felici was also modified through an efficient tyrosine residue phosphorylation using PTK $p55^{fyn}$ and became an important tool for studying *po*Y-binding modules. Tong and coleuses (Tong *et al.,* 2002) took advantage of this library for a systematic compilation of consensus motifs for 24 predicted yeast SH3-domains. This compilation was then used to derive a protein interaction network which links each recognition module to all those proteins found in the data base to contain a preferred peptide site (see also section 1.1.2.1).

Table 1.1.5: Screens of random peptides repertoires

Library	Domain	Protein	Ligand	Reference
M13 display at N-terminus of phage coat protein III				
GX_6G	SH3	HSrc	RSLPPIPG Table 1.1.6 entry 3	Rickles *et al.,* 1994
X_7, CX_7C	chromo-shadow	HP1	*fn(*P/L)(W/R/Y)VΦΦ*fc*, linear and constr.	Smothers *et al.,* 2000
CX_9C	SH2	Grb2	*fn*YxN*fc*, constr.	Oligino *et al.,* 1997
$X_{10}C$	EH	frog Intersectin	*Fn*NPF*fc*, constr.	Yamabhai *et al.,* 1998
$X_{10}C$		β_1-syntrophin	*fnx*[+]ETC(L/M)AgxΦC*fc*, constr.	Gee *et al.,* 1998
X_{12}	WW	hWWP1, hWWP2, hWWP3, hYAP, yRsp5.3, mNedd4	*fn*(L/P)PxY*fc*, class I, Table 1.1.6 entry 7	Kasanov *et al.,* 2001
X_{12}	SH3-1, SH3-2	CIN85	*fn*Px(P/A)xxR*fc*	Kurakin *et al.,* 2003
X_{15}	SH3	Src	*fn*RPLPxxP*fc*	Cheadle *et al.* 1994
$X_4CX_{10}CX_4$	SH2	Grb7	*fn*YxN*fc*, constr., (see Table 1.1.6 entry 8)	Pero *et al.,* 2002
$X_2CX_{14}CX_2$		h14-3-3τ	*fn*WLDL*fc*, constr.	Wang *et al.,* 1999
X_2CX_{18}		h14-3-3τ	*fn*RSx{1,3}[-]*fc*, linear, constr.	
X_{10}(G/C/R/S)G(V/A/D/G)-X_{10}[a)], X_{18}PGX_{18}[a)]	SH3	Src	*fn*RPLPPLP*fc*, (see Table 1.1.6 entry 2)	Sparks *et al.,* 1994

Library	Domain	Protein	Ligand		Reference
M13 display at N-terminus of phage coat protein VIII					
X_9	SH3	h amphiphisin h endophilin-2	*fn*RPxR*fc* *fn*P[+]RPPxPR*fc*		Cestra *et al.,* 1999
X_9	EH1-3	mEps15, mEps15R	*fn*NPF([+]/Φ)*fc* *fn*(N/H)(P/H)F[+]*fc*		Salcini *et al.,* 1997
X_9	EH1	mEps15	*fn*ζNPF([+]/Φ)*fc*	class I	Paoluzi *et al.,* 1998
	EH2		*fn*ζNPF([+]/Φ)*fc*	class I	
	EH3		*fn*FW*fc, fn*NPFW*fc*	class II	
	EH1	mEps15R	*fn*NPFR*fc*	class I	
	EH2		*fn*ζNPF([+]/Φ)*fc*	class I	
	EH3		*fn*(S/T)NPFRQ*fc*	class I	
	EH1	yYBL047C	*fn*Yxxx(F/L)WRP*fc,* *fn*NPF*fc*	class II class I	
	EH2		*fn*WWxxAD*fc* *fn*NPF*fc*	class II class I	
	EH3		*fn*Rxx*NPFR*fc*	class I	
	EH2	yPan1	*fnN*NPFxD*fc*	class I	
	EH1	yEnd3	*fn*SWGxxxW*fc* *fn*H(T/S)F*fc*	class II	
X_9	WW[b)]	Utrophin	*fn*PPxY*fc*		Di Vignano *et al.,* 2000
X_9	SH3	hEps8, mEps8, hEps8R2, mEps8R2	*fn*PxxDY*fc*		Mongiovi *et al.,* 1999
X_9	EH	frog Intersectin	*fn*NPF*fc*		Yamabhai *et al.,* 1998
X_9[c)]	PTB	Shc	*fn*F/YxNPT*po*YxxY/W*fc*		Dante *et al.,* 1997
X_9[c)]	SH2	Shc, Sli, Rai	*fn*N(I/V)*po*Y(E/G)T(I/V/L)(W/F)*fc* *fn*(L/I/V)*po*Y(E/G)T(W/Y/F)*fc* *fn*(V/I)*po*Y(E/G)(T/Y)(I/L)*fc*		Pelicci *et al.,* 1996
X_{10}, X_{12}, CX_8C	SH2	hGrb2	*fn*YxN*fc*, linear, constr. Table 1.1.6 entry 9		Hart *et al.,* 1999
M13 display at C-terminus of phage coat protein VIII					
X_7	PDZ2 PDZ3	MAGI3 kinase	*fn*(T/S)WV *fn*FDI	class I class II	Fuh *et al.,* 2000
λ phage display at C-terminus of the D protein					
X_9	PDZ1 PDZ2 PDZ3	hINDAL	*fn*(V/R)ΩWΦ *fn*EΩDF *fn*ΩDΦ,	class II class II class II	Vaccaro *et al.,* 2001

Library	Domain	Protein	Ligand		Reference
			*fn*Ω(D/E)	class IV	
	PDZ4		*fn*EΦΩV	class II	
	PDZ5		*fn*(S/T)W(V/L)	class I	
	PDZ6		*fn*ΦSΩV	class I	
	PDZ7		*fn*SxV	class I	
The T7 phage display at C-terminus of the 10B phage coat protein					
X_{16}	SH3-1, SH3-2 SH3-3	CIN85	*fn*Px(P/A)xxR*fc*		Kurakin *et al.,* 2003
Plasmid display					
X_{15}[d)]	PDZ3	PSD95	*fn*E(T/S)xV		Stricker *et al.,* 1997
X_{15}[d)]	PDZ	NNOS	*fn*DxV		

[a)] the non-X positions resulted from the cloning strategy used to generate the oligonucleotide library

[b)] interaction between peptides and WW domain occurred only in the presence of EF hands and ZZ domain

[c)] modified through an efficient tyrosine residue phosphorylation using PTK $p55^{fyn}$

[d)] peptides were fused to the C-terminus of the Escherichia coli lac repressor

1.1.4.2 Dedicated peptide libraries

The rather short and general consensus motifs identified by screening of the random libraries above often failed to explain the specificity of proteins pulled down from cell extracts (Rickles *et al.,* 1995). Specially constructed phage libraries displaying peptide ligand repertoires biased for particular domains have been constructed for a more detailed analysis of ligand preferences. Table 1.1.6 summarizes such experiments which were, in the majority, follow-up studies from experiments listed in Table 1.1.5. However, the predominance of such analyses was conducted using chemical peptide synthesis (see section 1.1.5). Schmitz *et al.,* 1996, prepared a peptide repertoire X_3YX_4 of eight amino acids at the N-terminus of M13 pIII to search for tyrosine kinase substrates. Following an enzymatic phosphorylation of the phage library, this repertoire was then suitable to study *po*Y binding modules. A similar library as that of entry 5 (Table 1.1.6), X_4PXXPX_4, was used by Tong *et al.,* 2002) in their systematic yeast SH3 domain study (see also section 1.1.2.1).

Table 1.1.6: Screens of oriented peptide libraries

Entry	Library	Domain	Protein	Ligand	Class[a)]	Reference
M13 display at N-terminus of phage coat protein III						
1	X_3YX_4[b)]	SH2	Grb2	*fnpo*Y(M/E)NW*fc*	II	Gram *et al.,* 1997
2	8 (aa)[c)]	SH3	Src	*fn*RPLPPLP*fc*	1 R	Sparks *et al.,* 1994
3	X_6PPIPG,	SH3	hSrc, avian	*fn*RPLPPLPxP*fc*	1 R	Sparks *et al.,* 1996
	RSLRPLX_6,		Src,	*fn*RPLPP(I/L)P*fc*	1 R	
	PPPYPPX_6		Fyn,	*fn*RxxRPLPPLPxP*fc*	1 R	
			Lyn,	*fn*RxxRPLPPLPPP*fc*	1 R	

Entry	Library	Domain	Protein	Ligand	Class[a)]	Reference
			P13K, MAbl	*fn*PPPYPPPP(I/V)P*fc*	1 @	
4	X_5RPLPPLPPP, RSLRPLPPLPX_5, GAAPPLPPRX_5	SH3	Src, Fyn, Lyn, P13K, Yes	for details see reference	1 R	Rickles *et al.,* 1995
5	X_6PXXPX_6	SH3	Src	*fn*LxxRPLPxΨP*fc*	1 R	Sparks *et al.,* 1996
			Yes	*fn*ΨxxRPLPxLP*fc*	1 R	
			Abl	*fn*PPxΩxPPPΨP*fc*	1 @	
			Cortactin	*fn*[+]PPΨPxKPxWL*fc*	2 K	
			p53bp2	*fnRPX*ΨPΨR[+]SxP*fc*	2 R	
			PLCγ	*fn*PPVPPRPxxTL*fc*	2 R	
			Crk	*fn*ΨPΨLPΨK*fc*	2 K	
			Grb2-N	*fn*[+]ΩxPLPxLP*fc*	I	
6	X_6PPX_6	WW	HYAP, mYAP	*fn*PPPPYP*fc*	I	Linn *et al.,* 1997
7	X_6PPX_6	WW	hWWP1, hWWP2, hWWP3, hYAP, yRsp5.3, mNedd4	*fn*(L/P)PxY*fc*	I	Kasanov *et al.,* 2001
8	X_4CX_4Y(D/A/E/X)-NX_3CX_4	SH2	Grb7	*fn*Y(D/A/E)N*fc*, constr	II	Pero *et al.,* 2002
M13 display at N-terminus of phage coat protein VIII						
9	X_5YXNX_8	SH2	HGrb2	for details see reference	II	Hart *et al.,* 1999

[a)] when available, sub-classification of SH3 modules is according to Cesareni *et al.,* 2002

[b)] modified through an efficient tyrosine residue phosphorylation using a mix of phosphotyrosine kinases Blk, c-Src and Syk (Rickles *et al.,* 1995)

[c)] library containing 8 amino acid long peptides encoded by the DNA sequence $((C/A)NN)_8$ which restricted the amino acid compositions to the following proportions: 6R:4P:4L:3I:2H:2Q:2K:2N:2S:1M

1.1.5 Synthetic peptide repertoires

Combinatorial and simultaneous parallel chemical synthesis techniques have been developed and refined since the early eighties and were applied very successfully to the preparation of peptide repertoires for studying molecular recognition events in immunology. This concerns the discovery and analysis of immunogenic sites (the “epitopes”) of protein antigens recognized by antibodies and T-cells (Liu *et al.,* 2003). Principally, the short functional sites recognized by the protein modules can be likewise considered as “linear epitopes”. Thus, the sophisticated approaches from immunology for library design and assay formats apply directly here.

Peptides are assembled through stepwise coupling of amino acid derivatives starting from the C-terminus tethered to a solid support material. There is a large variety of support materials and formats available with PEG-polystyrene (Rapp-Tentagel), polyethylene pins and cellulose membranes dominating the field. In addition to the 20 genetically coded aa, any non-natural synthetic derivative can be incorporated. Random positions in a library can be freely defined with respect to location and number in the sequence as well as the composition of aa residues, easily yielding theoretical diversities from a few hundred to several billion peptides. These positions are generated either by coupling a mixture of aa derivatives or by mixing together portions of support material from respective single aa coupling reactions ("mix and split"). Cysteines are often omitted or left protected (Acm-) to avoid random disulfide scrambling. After assembly and deprotection, peptides may be cleaved from the support and assayed in solution or are left tethered to the support for *in situ* solid phase binding assays. For more information see Ref. (Chan and White, 2000).

These synthetic opportunities are exploited to generate and screen peptide libraries in many formats (for a review see Frank, 2002), such as: A) selection from the full random pool, B) iterative or recursive screening by various sub-library pooling strategies with a combination of one or more defined aa positions in a random background, the best hit of one screening round is then used as the start for the next round; C) "positional scanning" of all consecutive positions by single aa incorporations, or "dual positional scanning" for two neighboring positions, respectively; D) orthogonal (self-deciphering) pools; E) epitope mapping (or "peptide walking"), the systematic scanning of a protein sequence with overlapping peptide fragments, F) replacement analysis, the systematic single substitution of each epitope residue by one mock aa (e.g. alanine scan); G) "replacement net" analysis, the systematic single substitution of each epitope residue by all other aa residues; H) "motif scanning", the systematic editing of an epitope sequence in an all alanine (A) or all random (X) background to define consensus recognition motives; I) finally, large collections of individual peptides derived by searching genomic sequences with epitope prediction algorithms (see section 1.2.1).

1.1.5.1 Soluble peptide libraries as ligand repertoires

Screening for module binding is achieved by an affinity chromatography selection with an immobilized domain/protein of interest. The eluted peptides are sequenced or analyzed by mass spectrometry as a mixture. This readily yields the consensus binding motif as result of the pool analysis. The random library, however, has to be oriented to one or more residues that are decisive for binding in order to avoid phase problems if the interacting residues on

different peptides were located at different positions. Table 1.1.7 summarizes some library screens reported in the literature.

Table 1.1.7: Screens of soluble peptide libraries.

Library	Specification	td	Screened domain	Protein of origin	Reference
$APXYSP_5$		20	SH3	Abl	Santamaria *et al.,* 2003
$APTXSP_5$		20	SH3	Abl	
$APTYSPXP_3$		20	SH3	Abl	
$MAX_4YX_4AK_3$	X – C,W	1.1×10^{10}	PTB	dNumb	Li *et al.,* 1997
***po*Y-oriented**					
$G(F_2Pmp)XPXS$-amide	X–C; F_2Pmp as mimic a poY	361	SH2	PI 3-kinase	Kelly *et al.,* 1996
GDG*po*YX_3SPL_3	X – C,W	5832	SH2	22 proteins	Songyang *et al.,* 1993
GAX_3*po*YX_3K_3	X – C,W	3.4×10^7	PTB	Shc, Cbl	Songyang *et al.,* 1995
MAX_3NXX*po*$YXAK_2$	X – C,W	3.4×10^7	PTB	ShcC	O'Bryan *et al.,* 1996
***poS/po*T-oriented**					
MAX_3*po*SX_3AKK	X – C	4.7×10^7	14-3-3	8 proteins	Yaffe *et al.,* 1997
MAX_4*po*SXPXXAKK	X – C	8.9×10^8	14-3-3	8 proteins	
MAX_3RXXR*po*SXPAKK	X – C	4.7×10^7	14-3-3	8 proteins	
$MAXXRXX$*po*SXPAKK	X – C	2.5×10^6	14-3-3	8 proteins	
MASX*po*SXXAKK	X – C	6.9×10^3	14-3-3	8 proteins	
ISRST*po*SX_3NK	X – C	6.9×10^3	BRCT	5 proteins	Rodriguez *et al.,* 2003
KAX_3*po*SX_3AK	X – C	4.7×10^7	BRCT	BRCA1	
MAX_4*po*TX_4AK_3	X – C,W	1.1×10^{10}	FHA	6 proteins	Durocher *et al.,* 2000
MAX_4*po*TXXIXXAK$_3$	X – C	1.7×10^{10}	FHA	3 proteins	
P-oriented					
$MAX_4PXXPX_3AK_3$	X – C,W	2×10^{11}	SH3	amphipysin	Grabs *et al.,* 1997
$MAX_4PXXPX_3AK_3$	X – C,W	2×10^{11}	WW	4 proteins	Bedford *et al.,* 2000
MAX_4PX_4AKK	X – C	1.7×10^{10}	WW	5 proteins	
$MAX_4PPRX_4AK_3$	X – C	1.7×10^{10}	WW	3 proteins	
$MAX_4YX_4AK_3$	X – C,Y	1.1×10^{10}	WW	YAP	
C-terminal-oriented					
KNX_8-COOH	X – C,W	1.1×10^{10}	PDZ	6 proteins	Songyang *et al.,* 1997
$KNX_6(S/T/Y)XX$-COOH	X – C,W	1.1×10^{10}	PDZ	5 proteins	

1.1.5.2 Bead-bound peptide libraries as ligand repertoires

A widely applied format of peptide libraries for bio-screening is the One-Bead-One-Compound (OBOC) combinatorial library format synthesized on resin beads by the mix and split technique (for a review see Lam *et al.,* 2003). The beads are of about 100 µm in size, carry about 100 pmol of a peptide sequence; one gram of resin contains about 3×10^6 beads. They are screened by incubation with a domain/protein of interest and bound protein is detected by various labels. The positive labeled beads are manually isolated with the help of a microscope and tweezers and are then individually sequenced by Edman degradation. Several examples of screenings performed using the OBOC combinatorial library format in context of various protein domains are described below. A biased combinatorial library of peptides having the following form; X_3PPXPXX-resin (X – C; td ~ 4.7×10^7) was synthesized and about 2×10^6 beads were screened with the fluorescently labeled SH3 domains of PI3K and Src (Yu *et al.,*1994). A combinatorial library AX_3*po*TX_3ABBRM-resin (X – C,M + norleucine as a substitute for M; td ~ 4.7×10^7) was screened with the biotinylated FHA1 and FHA2 domains of yeast Rad53 to determine their binding requirements (Liao *et al.,* 2000). A phosphotyrosine-oriented peptide library DEXX*po*YX_3IBBRM-resin (X – C,M + norleucine td = 2.5×10^6) was synthesised and about 2.86×10^6 beads were assayed with the N- and C-terminal SH2 domain of SHP-1 (Beebe *et al.,* 2000).

If much smaller beads (~10 µm) are used, rapid FACS technology can be exploited to separate fluorescent positive beads. However, the peptide material per bead is not anymore sufficient for sequencing and, thus, other strategies are required to identify the ligands. Pool sequencing of the sorted beads was applied to reveal the binding consensus of the Grb2 and Syk SH2 domain screened with a phosphotyrosine-oriented peptide library EPX^6*po*Y$X^{19}X^7X^{19}X^7X^6$ (the superscript numbers refer to the number of defined amino acids that were coupled at each position (td ~ 6.4×10^5); for a detailed matrix representation (see Muller *et al.,* 1996). A recursive deconvolution strategy was applied to investigate the Grb2 SH2 domain with a phosphotyrosine-oriented peptide library AS*po*YX_4SA (td = 1.6×10^5). First, twenty sub-libraries AS*po*Y$O^{+1}X_3$SA with one of the twenty amino acids fixed at the position O^{+1} were synthesized and incubated with a mixture of differently labeled target module and negative reference modules. The number of positive beads sorted for a particular aa/module combination was taken as a binding preference scale. In a series of synthesis and screening rounds, amino acid preferences for the consecutive random positions were determined likewise, taking the best hit of each previous screen. To select a high affinity motif the library was screened for beads that bind Grb2 but not the Grb2 R67H mutant. To search for the most selective ligand, Grb2 positive beads were sorted that do not bind to the SH2 domains of hAbl, Nck, bPI3Kp85-N, bPI3Kp85-C, mSHP-2N, vCrc and Grb2 R67H, applied as a negative reference mixture (Kessels *et al.,* 2002)

1.1.5.3 Peptide arrays as ligand repertoires

Synthetic peptides displayed at high density as an array on a planar support are a potent alternative tool for repertoire screening (for a recent review see Frank, 2002). Currently, the major technique for *in situ* synthesis and screening by this approach is the SPOT method. Individual peptide preparations are easily synthesized at different locations on a cellulose membrane by distribution of small (< µl) volumes of the respective solutions of activated aa derivatives or aa mixtures per spot and peptide elongation cycle (Frank and Owervin, 1996). Up to 2000 peptides or peptide pools can be synthesized on a single 8x12cm (MTP) size sheet of cellulose paper and probed simultaneously with a domain/protein of interest. Positive spots of the probed array immediately indicate the sequence of the peptide(s) recognized. Relative intensities of binding signals measured on a SPOT peptide array are to a good degree representative for relative complex affinities (Kd´s) as proven by Biacor measurements (Hultschig, 2000) and competitive ELISA (Kramer *et al*., 1999). SPOT synthesis is used to construct various types of libraries.

1.1.5.4 Sub-library pools for iterative a priori deconvolution

A library $XXO_aO_bSXV\text{-}COO^-$ with X (all combined 20 aa, except C, M and W) and O (all individual 20 natural aa) was synthesized as an array of the 400 sub-libraries for the O_aO_b combinations, each with a td of $17^3 = 4913$. This library was used to determine the aa preference of the syntrophin PDZ domain in the -3 and -4 position with respect to the essential C-terminal Val residue. For the display of a free carboxy terminus of the peptides, the tethering to the support had to be reoriented to the N-terminus! This was achieved first by coupling the N-terminus of the finished peptide back to the support in a cyclization reaction followed by cleavage of the C-terminal linker. Two descendant libraries $XOKESXV\text{-}COO^-$ and $XXKESOV\text{-}COO^-$ (each sub-library with td = 289) were synthesized to define requirements at –1 and –5 positions. Also heptamer sub-libraries of the type $X_4O_aXO_b\text{-}COO^-$, $X_5O_aO_b\text{-}COO^-$, $X_3O_aXXO_b\text{-}COO^-$, $X_3O_aXO_bX\text{-}COO^-$ and so forth (each sub-library with td ~ 1.4×10^6) were synthesized, but relevant binding affinity could not be detected (Schultz *et al*., 1998). An oriented library $AX_3(poS/poT)X_3A$ was used to evaluate binding specificity of four SH2 domains. The importance of different aa residues at any given position was estimated using a positional scanning library format. A total of 19 x 8 = 152 sub-libraries were spotted for the $AX_4(poS/poT)X_4A$ library, each sub-library had a td ~ 1.8×10^9. This method was designed as the “Oriented Peptide Array Library” (OPAL) Strategy (Rodriguez *et al*., 2004).

1.1.5.5 Protein Scanning Repertoires (Peptide walking)

With the increasing availability of genomic sequence information, and data from protein protein interactions, this approach is becoming more widely used. Libraries are characterized by “peptide length/offset” (offset is the number of aa residues by which overlapping peptide fragments are shifted along the protein sequence). A 15/6-scan of the *L. monocytogenes* ActA protein sequence (aa 30-614) was probed with radio-labeled VASP and led to the identification of the *fn*[-]FPPPPx[-]*fc* binding motif of the EVH1 domain (Niebuhr *et al.,* 1997). A 15/3 scan of the human zyxin sequence was probed with the EVH1 domain of Mena and homologous binding sites were found (Dress *et al.,* 2000). A 15/3-scan of the 140-kDa isoform of Mena revealed several binding sites for the FE65 WW domain all containing the PPLP sequence (Ermekowa *et al.,* 1997). A 15/1-scan of the C-terminal domain of mGluR5 (aa 1138-1169) identified the binding site LTPPSPF for the mVesl EVH1 domain, which recognizes the reverse of the Mena EVH1motif (Barzik *et al.,* 2001). The binding site PSIDRSTKP of the Gads C-terminal SH3 domain on SLP-76 (aa 181-291) was identified with a 15/4-scan (Berry *et al.,* 2002). A scan (11/2) of human synaptojanin (aa 1060-1312) was probed with the SH3 domains of amphiphysin and endophilin-2; several binding sites were identified, some matching the known consensus motifs and others do not (Cestra *et al.,* 1999).

1.1.5.6 Replacement repertoires

Once a peptide ligand is identified, the contribution of each single residue to the affinity and selectivity of the binding motif can be examined by systematic replacement approaches. Several examples are listed in Table 1.1.8

Table 1.1.8: Screens of replacement repertoires-

Protein-ligand	Binding site[a)]	Replacements	Domain	Protein	Reference
SLP-76	**PSIDRSTK**P	Alanine scan	SH3	Gads	Berry *et al.,* 2002
APP	**QNGYENPTYKFFEQMQN**	Alanine scan	PTB	Dab1	Howell *et al.,* 1999
Clone 13	**P_7LPAP_3QP**	Valine scan	WW	FE65	Ermekova *et al.,* 1997
Clone 13	P_7**L**PAP_3QP	Net	WW	FE65	Ermekova *et al.,* 1997
APP	QN**GY**E**NP**T**Y**KFFEQMQN	Net (except C)	PTB	Dab1	Howell *et al.,* 1999
ActA	SF**EFPPPPT**D	Net	EVH1	VASP	Niebuhr *et al.,* 1997

Protein-ligand	Binding site[a)]	Replacements	Domain	Protein	Reference
mGluR5	LEELVA**LTPPSPF**RD	Net	EVH1	mVesl	Brazik *et al.,* 2001
WBP-1	**GTPPPPYTVG**	Net	WW	hYAP	Chen *et al.,* 1997 Pires *et al.,* 2001
artificial	**CPK**PP**K**Y**PKK**	Net	WW	hYAP	Linn *et al.,* 1997
SLP-76	**PSIDRSTKP**	Net	SH3	Gads	Berry *et al.,* 2002
synaptojanin	**LPIRPSRAPSR**	Net	SH3	Amphiphisin endophilin 2	Cestra *et al.,* 1999
synaptojanin	**LEPKRP$_4$RP**	Net	SH3	amphiphisin endophilin 2	Cestra *et al.,* 1999
α-subunit the skeletal muscle VGSCs	**GVKESLV**	Net	PDZ	syntrophin	Schultz *et al.,* 1998
BACH1	**SRST*po*SPTFNK**	Net (except C)	BRCT	BRACA1	Rodriguez *et al.,* 2003
ActA	**SFEFPPPPTEDEL**	Net	EVH1	VASP	Ball *et al.,* 2000
ActA	**EFPPPPTEDELEII**	Net	EVH1	VASP	

[a)] aa positions replaced are in bold

1.1.5.7 Genome/proteome scanning

Literature and data base searches for functional sites of protein modules are an increasingly important and rich source for peptide ligands which can be synthesized and investigated using the SPOT method. A repertoire of 70 peptides that were described as SH3 or WW ligands was probed with VASP to assess the selectivity of the EVH1 domain (Niebuhr *et al.,* 1997). An array of all C-terminal heptapeptide sequences of human proteins found in Swiss-Prot release 34 (3514 sequences) was synthesized with a free carboxyl end by Hoffmüller *et al.,* 1999, and probed with the HRP-labeled PDZ domain of syntrophin. Landgraf *et al.,* 2004, took the binding motif cores of eight yeast SH3 domains that were obtained by phage display experiments (Tong *et al.,* 2002), to define for each one a less selective “relaxed pattern”. The *Saccharomyces* Genome Database was scanned for peptides matching these patterns. Approximately fifteen hundred peptides for each domain were selected for synthesis and probed with the corresponding SH3 domains. This procedure was called WISE (Whole Interactome Scanning Experiment). They reproduced this approach also with the SH3 domains of two human proteins, amphiphysin 1 and endophilin 1, for which peptide ligands were selected from the Swiss-Prot/TrEMBL database. A total of 3774 peptides were synthesized and tested with these two SH3 domains

1.1.5.8 Peptide arrays as domain repertoires

The WW domain, the smallest of all known protein modules, was the subject of a synthetic approach to study structure activity relation. A replacement net repertoire (cysteine excluded) of 837 individual variants of the 44 residue large hYAP WW domain was synthesized using the SPOT method. This array was probed with peroxidase-labeled EYPPYP$_4$YPSG peptide (Ball *et al*., 2000). Another repertoire of 11 859 variants of the hYAP WW domain was assembled by a combination of the SPOT method with native chemical peptide ligation. One set of 19^3 = 6859 WW domains (38-mers) with simultaneous substitutions at positions 30, 32 and 35 (making up the ligand binding pocket), by any of the 19 proteinogenic amino acids (excluding C). A second set of 5000 variants was synthesized bearing combinations of 19 proteinogenic, 20 non-proteinogenic and phosphorylated amino acids in positions 30, 32 and 35. These arrays were probed with 22 dye labeled different peptides ligands GTP$_4$XTVG (X–C but including *po*Y, *po*S and *po*T) (Toepert *et al*., 2003).

1.2 The experimental concept

All experimental approaches described above using peptide repertoires are more or less applied to analyze known families of domain interactions using peptide repertoires built around consensus binding motifs. There is missing a genome-wide search for new families and principles. This is the primary goal of the experimental concept underlying this thesis.

Both, Hawlisch *et al.,* 2001 and Bialek *et al.,* 2003 have shown that multiplex bio-panning with phage-displayed protein libraries can be carried out with a multitude of peptide targets arrayed on a planar surface. The project, thus, aims to utilize the two-dimensional screening concept of a library of peptide fragments versus a library of protein domains; both libraries should include all relevant components from a biological model system of interest such as a microorganism, eukaryotic cell or tissue. Both libraries, thus, are more or less genome spanning and considerably large. Peptide fragments will be displayed as immobilized synthetic compounds in the format of macroarrays. These are manufactured on planar continuous supports by in situ parallel chemical synthesis utilizing the SPOT technique (see section 1.2.2). A bacteriophage library displaying protein fragments expressed from a randomly fragmented cDNA library cloned into the phage genome to yield fusions with an envelope protein will be the source for the protein domains. The screening through large numbers of peptides simultaneously (multiplexed) should be possible with only a minimum of panning rounds, preferably with only one, to make this approach practicable. Furthermore, there will be more than just one, rather a family of protein domains enriched on one peptide ligand; members of this family will have different affinities toward the peptide ligand. Obviously, this situation requires a very competent analytical tool which can identify the

multitude of specifically enriched phages in a huge background of unspecific phages and all this in a high-throughput process.

1.2.1 Accessible Surface Scanner

Proteins interact via surface accessible interaction sites which involve amino acid side chain and backbone contacts along a linear segment of the protein chain (linear/continuous epitopes) or involve amino acid residues from two or more segments of the protein chain brought together by the folded conformation (conformational epitopes). Continuous binding sites can be efficiently represented by synthetic peptide fragments (Frank, 2002). Several possibilities to generate diverse peptide repertoires were briefly described in the section 1.1.5. When taking into account that continues epitopes correspond to the surface exposed regions of proteins and a large part of a protein determines its core (Atassi, 1979; Atassi and Kuriaski, 1984), the synthesis of only the exposed regions that contain potential interaction motifs seems reasonable and will reduce the effort of the approach considerably (see Fig 1.2.1).

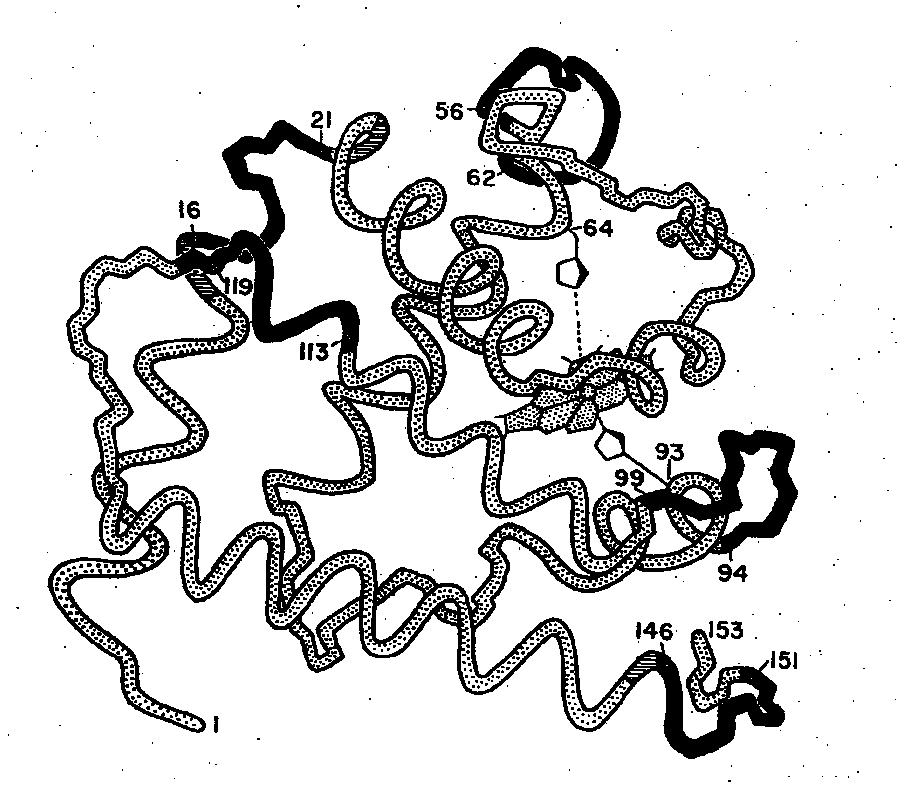


Fig 1.2.1: Drawing showing bio-active “epitopes” exposed on the protein surface (in black). Taken from Atassi, 1979.

One of the approaches to identify exposed regions without knowing the 3D-strucure involves computational scanning of the genomic sequences using epitope predicting algorithms. Several algorithms based on hydrophilicity (Kyte and Doolittle, 1982), chain flexibility (Karplus and Schultz, 1985), turn prediction (Chou and Fasman., 1978), solvent accessibility (Rost, 1996) and surface probability (Emini *et al.,* 1985) have been proposed for prediction of the protein exposed regions. Some of these (surface accessibility, chain flexibility and turn prediction) have been combined in a computer program, which predicts antigenic index (Jameson and Wolf, 1988), claimed to be more powerful than any single predictive parameter. For predictions of surface exposed protein regions from genomic sequences and convenient generation of peptide repertoires, a software tool, called Accessible Surface Scanner (ASS) was developed, which is entirely based on the Emini's surface probability algorithm. Fig 1.2.2 shows interface of the ASS software tool.

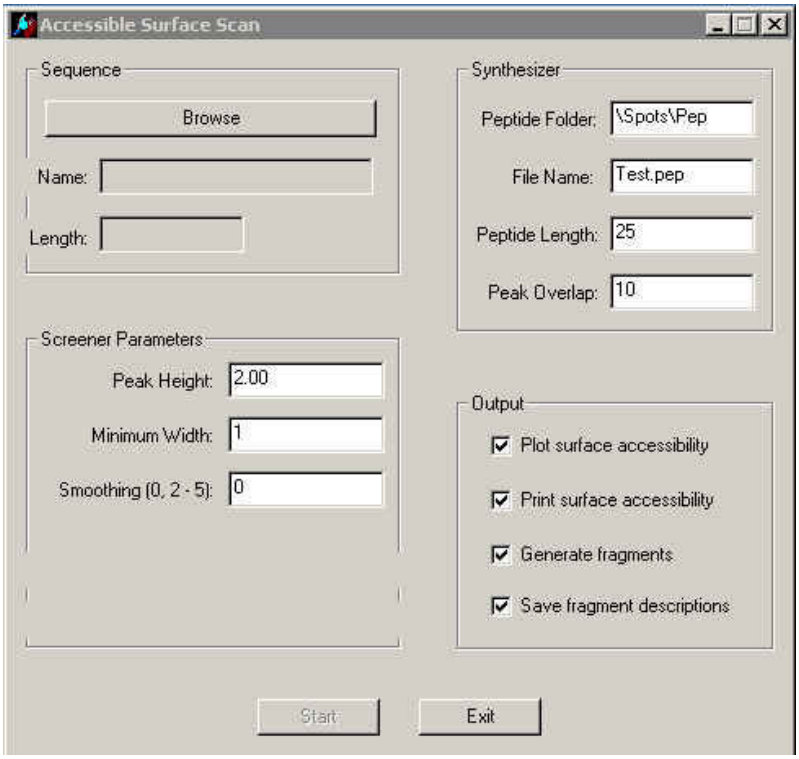


Fig 1.2.2: Picture showing interface of the Accessible Surface Scanner software (ASS, by David Lincoln)

The algorithm utilizes the empirical amino acid accessible surface probabilities of Janin which are fractional probabilities (0.26 to 0.97) determined for an amino acid found on the surface of a protein, based upon structural data from 28 proteins (Janin *et al.,* 1978). For a given amino acid sequence, a surface probability for amino acid "n" is a normalized product

of the surface probabilities of amino acids in positions n-2 to n+3. For detailed information refer to Emini *et al.,* 1985. The surface probability for a random hexapeptide is equal 1.0. The probabilities greater than 1.0 indicate an increased probability for the hexapeptide of being found on the surface. Large repertoires of peptides can be generated using the ASS software, as unlimited number of protein sequences can be scanned simultaneously. Protein sequences can be downloaded from genomic and proteomic databases in “fasta” format. Software parameters can be easily varied, which allows for more or less stringent surface probability predictions. Different length of generated peptide can be defined by typing desired number in peptide length window. Peak overlap parameter can be varied in a similar way. For example, in the case of setup shown in Fig 1.2.2, an appropriate number of 25 aa peptides overlapping with 10 aa will be generated from a surface probable fragment that is longer than the defined peptide length. Peak height parameter, defines threshold of surface probability. Amino acids with higher surface probability values than the defined threshold are considered as surface probable and are proposed peptide fragments to be synthesized. Minimum width parameter defines number of amino acids with values greater than the peak height threshold that have to be next to each other in order to qualify a particular region for peptide generation. Next parameter was introduced for smoothing of surface probability values using Savitzky and Golay filters. For detail refer to Savitzky and Golay, 1964. Following scanning of multiple protein sequences, resulting output files contain generated peptides only. In the case of calculations performed for a single protein sequence, a surface accessibility plot and a list of all amino acid residues with their surface accessibility values are available also. Advantageously, output peptide files are fully compatible with software controlling the automated peptide SPOT synthesis, which allows for convenient and safe transfer of data and synthesis of peptide repertoires.

1.2.2 The SPOT synthesis technique

Arrays of peptides for multiplexed phage-panning will be generated with the SPOT-synthesis method. SPOT synthesis belongs to the family of positional addressable solid phase methods. It is an easy and very flexible technique for simultaneous parallel chemical synthesis on membrane supports (Frank, 1992). In principle, dispensing of a small droplet of liquid on a planar surface of a porous membrane results in its absorption and formation of a circular spot. When appropriate reagents, dissolved in a low volatility solvent are spotted on defined positions, such spots form open reactors for chemical conversions involving the reactive functions anchored to the membrane support, e.g. conventional solid phase synthesis (Merriefield, 1963). The SPOT technique is especially widely used for generation of large repertoires of peptide fragments in a high-density array format. In this case, low

volatility solutions of N-terminally and side chain protected amino acids are distributed by manual or automated pipetting to defined positions on a modified cellulose membrane and are coupled in situ to the previously functionalized solid support see Fig 1.2.3. To prepare the membrane supports for chemical peptide synthesis the entire cellulose membranes were originally modified by esterification β-alanine (Fmoc-β-Ala) followed by removal of Fmoc protection group. Nowadays, specially modified membranes for preparations of either immobilized or solution-phase peptides are available commercially (see section 2.2.6). Novel types of special safety-catch linkers have been developed to cleave also C-terminally unmodified peptide acids or amides directly into neutral aqueous buffers for direct use in bioassays. Fmoc-β-Ala is initially distributed on the previously defined positions on the membrane generating sites for peptide synthesis. The remaining amino groups on the membrane are acylated and therefore no longer accessible for further reactions. Elongation of peptide sequence is an iterative process employing several washing steps as well as steps of de-protection, coupling and acylation of amino acids. The technical description of the entire process is in the material and method chapter (see section 2.2.6). In general, the N-terminal protection groups are removed before distribution of different activated amino acids to different peptide sites. After coupling reaction is completed, the remaining free amino functions are deactivated by acylation. In the next step the protecting groups of coupled amino acids are removed and the entire cycle can be repeated. After coupling the last amino acid, all peptides are N-terminally acetylated to mimic the backbone of proteins.

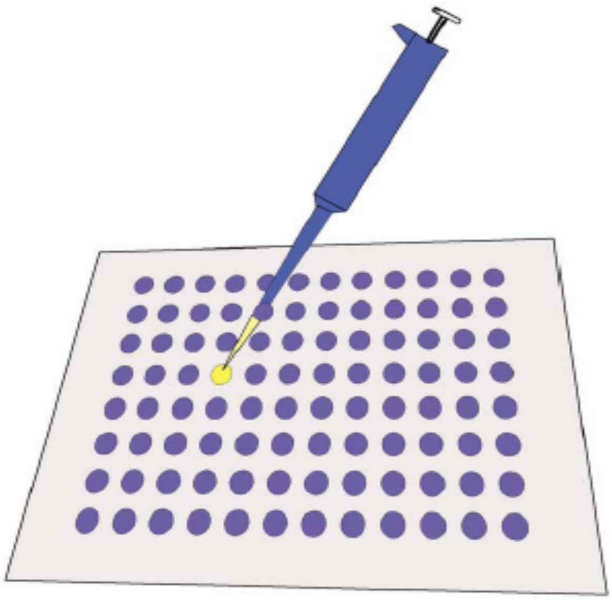


Fig 1.2.3: Schematic illustration of the principle of the SPOT method.

Despite several steps during SPOT peptide synthesis must be carried out manually, the delivery of the activated amino acids to corresponding reaction sites has been automized using the SPOT synthesizer (Fig 1.2.4) developed at ABIMED Analysen-Technik GmbH (Langenfeld, Germany). The process automation helped to reduce the size of spots and thus, increase the number per area considerably. Currently, 2500 spots can be generated on a microtiter plate size sheet. This is definitely helpful in a high-throughput application of this technique. SPOT peptide arrays are particularly suited for in situ biological activity screening in binding or enzymatic transformation assays. They became a very important tool for studying of e.g. protein-protein/peptide interactions, mapping of immunological epitopes, studying of binding motifs of protein modules. Several examples of application of SPOT-peptide arrays for different approaches concerned with protein-protein interactions were shown in the previous section. Much wider scope of their applications is reviewed in Frank, 2002.

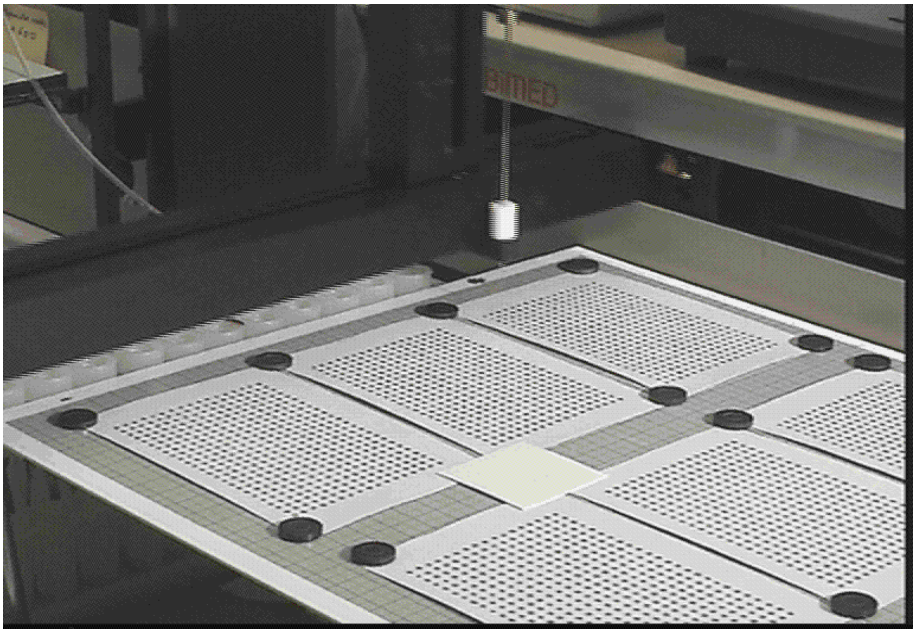


Fig 1.2.4: Picture showing the ABIMED peptide synthesizer delivering activated amino acids on the appropriate reaction sites.

1.2.3 T7 and lambda phage display systems

Several phage display systems are available for the anticipated approach, for a recent review refer to Castagnoli *et al.,* 2001. Three most popular phage display systems are based on filamentous (e.g. M13) and virulent T7 and lambda bacteriophages. The M13 phage has been successfully used for display of peptides (see section 1.1.4) and antibodies or their

fragments (Clackson *et al.,* 1991; Hoogenboom *et al.,* 1998). Serious limitations of this system derive from the characteristics of the phage life cycle. Filamentous phages are released from the host cell without breaking the integrity of the cell membrane, and proteins that are eventually assembled to form the capsid must cross the lipid bilayer of the inner membrane. Thus, any peptide or protein whose properties prevent it from crossing the membrane will stop the correct transfer of the hybrid capsid protein which will not be assembled. Finally, other limitations derive from the chemical characteristics of the periplasmic environment which may affect the folding and stability of the protein to be displayed (Castagnoli *et al.,* 2003). Contrary to this, virulent phages assemble in the cytoplasm and lyse bacterial membrane. This is advantageous for display of large proteins, which can not be efficiently transported through the lipid bilayer of. Therefore virulent bacteriophages seem advantageous for display of various genomic protein libraries and are more suitable for genome wide screenings purposed in this work. Several significant differences in surface display of proteins were reported for both phage types. This are mainly concerned with the number and size of proteins presented as fusions with capsid proteins (Castagnoli *et al.,* 2001). Comparison of display efficiencies for both T7 and lambda virulent phages is summarized in Table 1.2.1.

Table 1.2.1: Main characteristics of the virulent T7 and lambda display systems. Maximum size refers to the size of the heterologous protein successfully fused to the coat protein indicated. Castagnoli *et al.,* 2001

	T7			**Lambda**			
Protein	p10		pV	pD			
Copy/capsid	415		192	405			
C/N terminal	C		C	N		C	
Density	High	Low	Low	High	Low	High	Low
Maximum size (aa)	<50	<900	>1024	ND	286	>300	1024

Both display system presents very similar characteristics with respect to efficient display of short peptides at high density. Nevertheless, the gpD capsid protein of lambda phage seems more tolerant towards large (>300 aa) proteins displayed at high copy number. Higher display efficiency of lambda phage, increases avidity of displayed proteins, which is advantageous for selection of weaker interacting candidates of large size. On the other hand, the T7 bacteriophage has a shorter life cycle and, as a consequence, the time needed to perform multiple panning cycles is substantially decreased. Furthermore, differently from lambda, the T7 phage particle is extremely robust and resistant to harsh conditions that can be used to elute hybrid phage particles bound to their ligands without affecting phage

infectivity (Castagnoli *et al.,* 2001). Large diversity of T7 phage display cDNA libraries (T7Select) achieved from different, diseased or healthy tissues of different organisms are commercially available from Novagen. Libraries utilizing lambda phages (λgt10 and λgt11) can be purchased from Clontech. Protein and peptide libraries displayed on lambda phage were also contracted by other researchers (Moriki *et al.,* 1999; Santi *et al.,* 2000; Zucconi *et al.,* 2001; Sternberg and Hoess, 1995).

1.3 Aim of this thesis

The general goal will be to develop a process for the genome-wide interaction mapping between protein domains and peptide ligands based on the experimental concept described above and which is robust and truly high throughput. To realize this, several different tasks will have to be solved, briefly described below.

As mentioned before, the project is based on two dimensional screening of genome-complement repertoires of protein domains displayed on phage particles versus SPOT peptide-ligands arrayed on a planar surface. Before employing such large repertoires, a set of three model protein domains and their respective, known proline-rich peptide-ligands will be applied to evaluate technical prerequisites of the concept. Table 1.3.1 lists the properties of these domains.

Table 1.3.1: Characteristics of the three model protein domains selected for proof-of-principle studies

Code	Type of domain	Parent protein	Peptide ligand	*K*d (μM)	DNA size (bp)[a]	Domain size (aa)	Reference
E	EVH1	mMena	SFEFPPPPTDEELRL-pep-E	5	467	113	Niebuhr *et al.,* 1997
Y	WW	hYAP	GTPPPPYTVG-pep-Y	54	278	64	Chen and Sudol 1995
F	WW	RFE65	PPPPPPPLPAPPPQP-pep-F	n.d.	228	46	Ermekova *et al.,* 1997

[a] Size of the PCR product with TS21 and TA20 primers (see Table 2.1.6.1 in section 2.1.6)

The most important features of the model domains for this study are their different affinities for the peptide ligands covering lower and higher affinity interactions as well as their characteristic difference in the size of the coding DNA to allow for an easy identification by PCR and gel electrophoretic sizing. Previously, these model domains were cloned into the T7 phage chromosome to be displayed on the phage surface and successfully applied for affinity selections on their peptides on spots. Similarly, three other, indifferent proteins Fas

antigen (Siegel *et al.,* 2000), Ezrin (Gary and Bretscher, 1995) and Interleukin-4 (Singh, 2003), were cloned in he same chromosome and used for dilution of model domains in selection experiments. These studies were carried out in order to set up the bio-panning process and phage recovery from the spots (for details see Bialek *et al.,* 2003).

Here, the model domains will be cloned also into the chromosome of lambda phage and these two different phage display systems will be compared with respect to peptide-specific enrichment on peptide arrays containing respective ligands. The major difference between these phage vehicles is the density of protein presentation on the phage capsids, which is considerably higher for lambda, as described in section 1.2.3. This will result in a higher avidity of lambda complexes with the high local density of immobilized peptide ligands on spots of the planar surface. It was also reported that the selectivity for high-affinity ligands is best with T7 and for low-affinity ligands best with lambda (Castagnoli *et al.,* 2001). Because of the differences in binding affinities between the model domains, they are perfect for such type of comparison. This will help to chose better performing phage display system for future applications.

Another task will be concerned with an evaluation of the ASS computer software tool described in section 1.2.1, which was developed for prediction of surface-exposed peptide fragments from protein sequence data. A set of sequences of proteins containing known binding motifs for different protein domains will be scanned using different parameters. Additionally, computational surface probability predictions will be compared with crystal structures of analyzed proteins. The most optimal parameters set deduced from these studies will be employed for generation of large peptide repertoires from sequences of native proteins. Peptides will be synthesized in the form of spot macroarrays on the cellulose membranes. Resulting peptide macroarrays will be used for high-throughput phage panning experiments with protein domain repertoire, employing experimentally selected display system.

Obviously, the most important task in this thesis will be the establishment of a robust and practicable analytical method allowing for identification and validation of protein domain-peptide interactions. This method must be competent and sensitive enough to distinguish between many different specific and non-specific binding events after only one cycle of affinity selection.

Finally, for purpose of fitting all elements into an automated high-throughput process, affinity enrichment tests on a novel structurally modified plastic surface, which is suitable for spatially addressable parallel peptide synthesis and phage elution, will be performed. This surface is designed to prevent contact of liquids added to different neighbouring peptide spots for dissociation of bound phage populations.

2. Material and methods

2.1 Material

2.1.1 Chemicals/reagents, materials, kits and enzymes

All chemicals used in the context of this thesis were of analytical quality. Aqueous solutions were prepared with double deionized (Milli-Q) water. Where applicable, solutions were autoclaved. All chemicals, materials, kits and enzymes that are not shown in Table 2.1.1.1 were purchased from Aldrich, Amersham Pharmacia, Difco, Fluka, GibcoBRL, Merck, Millipore, New England Biolabs, Novagen, QIAGEN, Riedel de Haen, Roche, Roth, Sigma, Schmidt, Greiner, AB Gene and Stratagene.
Fmoc derivatives of amino acids and further reagents for peptide synthesis were obtained from Bachem, Calbiochem and Alexis.

Table 2.1.1.1: Chemicals/reagents, materials, kits and enzymes

Product	Provider
Chemicals/Reagents	
Albumin bovine serum	Sigma; Steinheim, Germany
Betaine, PCR Reagent	
Heparin	
MTT	
Alkaline Phosphatase-conjugated Streptavidin	Dianova; Hannover, Germany
BCIP	Biomol GmbH; Hamburg, Germany
IPTG	C.Roth GmbH; Karlsruhe, Germany
L.M.P Agarose	Invitrogen GmbH; Karlsruhe, Germany
peq GOLD Universal Agarose	peqLab Biotechnologie GmbH; Erlangen, Germany
Skim Milk (Difco)	Otto Nordwald KG; Hamburg, Germany
Proteinase inhibitor cocktail	Sigma; Steinheim, Germany
DNA MB grade from fish sperm	Roche Diagnostics GmbH; Mannheim, Germany
dNTPs	New England Biolabs GmbH; Frankfurt am Main, Germany
Cyanine 5-dCTP	Perkin Elmer Life sciences, Inc; Boston, MA,
Cyanine 3-dCTP	USA

Material	
Amino-PEG-Cellulose membranes	AIMS Scientific Products GmbH; Braunschweig, Germany.
Protran BA 85 membrane 0.45μm, ϕ 132 mm	Schleicher & Schuell; Dassel, Germany
Amine Silinated Slides	Cell Associates Inc; Pearland, Texas, USA
Super Amine Substrates	TeleChem International Inc; Sunnyvale, CA, USA
MultiScreen – PCR Plates	Millipore; Eschborn, Germany
96 well PCR plates Thermo-Fast 96, Skirted	AB Gene; Epsom, Surrey, UK
384 well microarraying plates	
Adhesive PCR film	
Gas permeable adhesive seals	
Square 96 well blocks	Qiagen; Hilden, Germany
Dialysis membrane 12 – 14 kDa	Meticell International Ltd, London, UK
Kits	
QIAprep Spin Miniprep Plasmid Kit	Qiagen; Hilden, Germany
QIAquick Gel Extraction Kit	
PCR Purification Kit	
R.E.A.L Prep 96 plasmid kit	
Montage 96 plasmid miniprep kit	Millipore; Eschborn, Germany
Nucleobond AX 100 kit	Macherey-Nagel; Düren, Germany
ReadyTo Go Lambda Packaging Kit	Amersham Pharmacia; Freiburg, Germany
Enzymes	
Shrimps alkaline phosphatase	USB cooperation, Cleveland, USA
*Hind*III	New England Biolabs GmbH; Frankfurt am
*EcoR*I	Main,
*Not*I	Germany
*Spe*I	
*Xho*I	
Taq DNA Polymerase	
Taq DNA Polymerase	Promega GmbH; Mannheim, Germany
T4 DNA Ligase	
HotGoldstar DNA Polymerase	Eurogentec Deutschland GmbH; Köln, Germany
Pfu polymerase	MBI Fermentas GmbH, St. Leon-Rot, Germany
Self-made (SM) *Taq* DNA Polymerase	GBF, Braunschweig (see 2.2.1)
Antibodies	
rabbit anti-Mena PAK 73	Purchased from Dr. Melanie Barzik
Goat anti-rabbit IgG, POD-conjugated	Jackson Immuno-Res, Cambridgeshire, UK
Rabbit anti-GST, POD-conjugated	Sigma; Steinheim, Germany

2.1.2 Devices

Table 2.1.2.1: Devices used for this work

Device	Symbol/Description	Provider
SPOT synthesis		
Automated SPOT robot	Abimed ASP222	Abimed Analysen Technik; Langenfeld, Germany
Microarray		
Contact printing robot	MicroGrid II TAS	BioRobotics; Cambridge, UK
Printing Split Pins	Microspot 2500	GeneWorx AG; Oberhaching, Germany
Hybridization station	Lucidea Slide Pro	Amersham Biosciences; Buckinghamshire, UK
Biochip Reader	ArrayWorx	Applied Precision; Issaquah, WA, USA
Biochip Reader	GMS 418	Genetic Microsystems; Woburn, MA, USA
Agarose gel electrophoresis		
Agarose gel tanks	Agagel Mini, Agagel G45, Agagel Maxi	Biometra Biomedizinische Analytik GmbH; Göttingen, Germany
Agarose gel tank	Electro-Fast WIDE SYSTEM	AB Gene; Epsom, Surrey, UK
Power supply		
High Voltage Power Pack	P30	Biometra Biomedizinische Analytik GmbH; Göttingen, Germany
Minicell Power Pack	P040-000	
Gel documentation		
CCD Camera controlled by EASY Image 3.16 software		Herolab; Wiesloch, Germany
UV Transiluminator	UVT-20	
PCR devices		
MJ Research Peltier Thermal Cycler	PTC 200	Biozym Diagnostics GmbH; Hess. Oldendorf, Germany
MJ Research Thermal Cycler	Tetrad with 4 alpha blocks	
Centrifuges and rotors		
Sorvall centifuge	RC5C	Kendro Laboratory Products; Bad Homburg, Germany

IEC centrifuges	Micromax RF Centra-GP8R	Int. Equip Company/Life Sciences International; Frankfurt, Germany
Sorval rotors	SS-34 SLA-1500 SLA-3000	Kendro Laboratory Products; Bad Homburg, Germany
Photometers		
RNA/DNA calculator UV/Visible Spectrophotometer	GeneQuant II Ultrospec 2000	Pharmacia Biotech Europe GmbH; Freiburg, Germany
Other equipment		
UV Crosslinker	Stratalinker 1800	Stratagene; La Jolla, CA, USA
pH-Meter	CG 840	Schott-Geräte GmbH, Hofheim, Germany
Thermomixer	54 36	Eppendorf; Hamburg, Germany
Incubator	Multitron	Infros HT GmbH; Einsbach, Germany
Shakers	Rotamax 120 Duomax 1030	Heidolph Instruments GmbH & Co. KG; Schwabach, Germany
French Press	French® Pressure Cell Press	SLM-Aminco; Rochester, USA
Electroporator	Gene Pulser II	Biorad Laboratories GmbH; München, Germany
Multichannel Electronical Pipet	Finnpipette BioControl	Labsystems; Helsinki, Finland
Evaporator centrifuge	Univapo 150 H	UniEquip; Martinsreid, Germany
Vacuum Manifold	MultiScreen Separation System	Millipore GmbH; Eschborn, Germany

2.1.3 Computer software

Several computer programs were used in the context of this work. Accessible Surface Scanner software version 1.0 was used for selection of peptides for SPOT synthesis (developed by David Lincoln). Microarray analysis was made with ImaGene 5.5.2 and GeneSight 3.5.2 software (BioDiscovery, Inc; Segundo, USA). TAS Application Suite 2.2.2.8 (BioRobotics; Cambridge, UK) was explored for set-up parameters and printing of microarray cDNA biochips. A hybridization protocol was developed with Lucidea Slide Pro 2.0 software (Amersham Biosciences; Buckinghamshire, UK). Live Image 1.51.0.42 (Genetic

Microsystems; Woburn, MA, USA) and ArrayWorx 2.1.2 software suite (Applied Precision; Issaquah, WA, USA) were used for scanning of the microarray biochips. DNA sequencing data were evaluated with Segman II a Lasergene module (DNA Star, Inc; Madison WI, USA). Oligonucleotide primers were engineered with Vector NTI 3.1 (InfroMax, Inc; Gaitherburg, MD, USA). Images were viewed and converted with Jasc Paint Shop Pro 4.1.

2.1.4 Bacteria, bacteriophages and clone libraries

Different *E. coli* strains used in this work are shown in Table 2.1.4.1. Bacteriophage lambda, λDisplay1 was a kind gift from Prof. Gianni Cesareni, the University of Rome, Italy. Adult Human Brain phage-display library, T7select10-3b was purchased from Novagen; Madison, USA. The full-length cDNA clone library of human brain amygdala region as well as two 384 well clone plates selected from the other two libraries were purchased from RZPD; Berlin, Germany.

The amygdala library: Name Hamy 2; No. 761, Plates No. 1, 2, 7 – 20; 6144 clones

Table 2.1.4.1: *E. coli* strains.

***E. coli* - strain**	**Genotype/Phenotype**	**References/Provider**
BB4	3,4 LE392.23 [F´ *lacI*qZ∆*M15 proAB* Tn*10* (Tet^R)]	Greener, 1993 Bullock *et al.,* 1987 Stratagene
BLT5403	F^- *ompT hsdS*B (rB^- mB^-) *gal dcm* pAR5403 (Ap^R)	Novagen
BLT5615	F^- *ompT hsdS*B (rB^- mB^-) *gal dcm* pAR5615 (Ap^R)	Novagen
DH10B™	F^- *mcr*A (*mrr-hsd*RMS-*mcr*BC) 80*lac*Z M15 *lac*X74 *rec*A1 *end*A1 *ara* 139 (*ara*, *leu*) 7697 *gal*U *gal*K - *rps*L (Str^R) *nup*G	Invitrogen Grant *et al.,* 1990
Gene Hogs, DH10B™-derived *E. coli* strain	F^- *mcr*A *mrr-hsd*RMS-*mcr*BC) 80*lac*Z M15 *lac*X74 *rec*A1 *end*A1 *ara*D139 (*ara-leu*) 7697 *gal*U *gal*K *rps*L (Str^R) *nup*G	Invitrogen
BL21(DE3)pLysS	F^- *ompT hsdS*B (rB^- mB^-) *gal dcm* (DE3) pLysS (Cm^R)	Studier *et al*., 1990 Novagen

2.1.5 Vectors

All vectors used in this work are briefly described in Table 2.1.5.1. Plasmids/phagemids were isolated from library clones obtained from RZPD.

Table 2.1.5.1: Plasmids, phagemids and recombinant vectors engineered for the purpose of this work.

Plasmid/phagemid	Genotype/Phenotype	References/Provider
pSPORT 1	General: Phagemid ds-DNA 4019 BP Function: cloning, sequencing Promoter: SP6, T7 Resistance: amp Ori: pUC Host: *E. coli* Related to: pSPORT series	Available from Invitrogen, as a clone from RZPD
pCMV-SPORT 6	General: Phagemid ds-DNA 4396 BP Function: cloning, eukaryotic expression Promoter: SP6, T7 Resistance: amp Ori: pUC, SV40 Host: *E. coli*, mammalian Related to: pCMV-SPORT series	Available from Invitrogen clone source: RZPD Berlin
pOTB 7	General: Plasmid ds-DNA 1815 BP Function: cloning Promoter: SP6, T7 Resistance: cm Ori: pUC Host: *E. coli* Related to: pOTB 2	Link: www.fruitfly.org Available as a clone from RZPD
Phage display vectors		
T7Select10-3b	T7Select series	Novagen; Rosenberg et. al,1996
λDisplay1	Related to: λpRH825 and λ171	Castagnoli *et. al,* 2001 Santi *et. al*, 2000
Recombinant vectors engineered for purpose of this work		
pCMV-SPORT 6 + E, F, Y, Fas, IL4, Ezr	Refere to pCMV-SPORT6	This thesis
pOTB 7 + E, F, Y, Fas, IL4, Ezr	Refere to pOTB7	This thesis
λDisplay1 + E, F, Y	Refere to: λDisplay1	This thesis (3.1.1)

2.1.6 Oligonucleotides

Oligonucleotides were purchased as a lyophilised DNA from: Invitrogen GmbH; Karlsruhe, Germany, MWG; Ebersberg, Germany and Biospring; Frankfurt am Main, Germany. Stock solutions at a final concentration of 100 pmol/ml and 10 pmol/ml were prepared by resuspension of lyophilised DNA in autoclaved Milli-Q water or TE-buffer. Aliquots were stored at –20°C. Oligonucleotides shown in Table 2.1.6.1 were used for PCR, cloning and sequencing.

Table 2.1.6.1: Oligonucleotides: Recognition sites of endonucleases are underlined

Primer pair	Sequence (5' → 3')
Protein domain specific primers for PCR amplification of DNAs coding for Mena-EVH1, WW/FE65, WW/Yap domains and following proteins: Fas antigen, IL-4, Ezrin	
E1	CGTGAATTCCATGGCTGAACAGAGT
E2	GGTAAGCTTCGACGTAGATCCTGTCAAT
F1	TCTGAATTCCGATCTACCGGCTGGA
F2	ATGAAGCTTCCCCTGTGATGGGGAG
Y1	GGCGAATTCTTCTTTTGAGATACCTGATGA
Y2	AATAAGCTTCGACTGGTGGGGGCTG
Fas S	TACGAATTCAGGCCGCCCGCTGTTTTC
Fas A	ACTAAGCTTCCATCATGGGTGGCAGGCTCT C
IL S	ACGGAATTCAGCTATTGATGGGTCTC
IL A	GCAAAGCTTAAAGCATGGTGGCTCAGTA
Ez S	CCCGAATTCTACAGCCGAATAGCCGAGG
Ez A	GTAGAAGCTTGGCCCCGTAACTGCTCTTTAAGGGACTGG
Primers used for PCR amplification and sequencing of T7Select 10-3b cDNA library clones	
TS 21	GGAGCTGTCGTATTCCAGTCA
TA 20	AACCCCTCAAGACCCGTTTA
Primer used for transfer of DNA inserts from T7Select 10-3b vector into λDisplay1 vector	
T7S	TCAACTAGTATGCTCGGGGATCCGAA ***SpeI***
T7A	CGTTACCTAGTTACTCGAGTGCGGC
Primers used for PCR amplification of DNA inserts cloned into λDisplay1 vector	
LS	CGGCAATCAGCATCGTTACT
LA	CGAATTCCTTAGCGGCCG
Primers used for PCR amplification of DNA inserts cloned into pCMV-SPORT 6 vector	
RZPD T7	TAATACGACTCACTATAGGG
RZPD SP6	ATTTAGGTGACACTATAG
Primers used for PCR amplification of DNA inserts cloned into pOTB 7 vector	
M13 F (-21)	TGTAAAACGACGGCCAGT

M13 R (19)	GGAAACAGCTATGACCATG
Primers used for PCR amplification of DNA inserts cloned into the pSPORT1 vector (full length cDNA amygdala library)	
M13 for 23	GACGTTGTAAAACGACGGCCAGT
M13 rev 30	ATAACAATTTCACACAGGAAACAGCTATGA
Random 15-mer primers labelled with Cy3 dye used for quality check of the microarrays	
Cy3Rand15	Cy3 (N) 15

2.1.7 DNA-Molecular Weight Markers

DNA standards most frequently used in this work are shown in Table 2.1.7.1. Depending on the agarose gel pocket sizes, different amounts of a particular standard were used and will be shown where applicable.

Table 2.1.7.1: DNA-Molecular weight markers.

Name	Mass range	Provider
Smart Ladder	200 – 10.000 bp	Eurogentec GmbH; Köln, Germany
MWM XIII	50 – 2642 bp	Boehringer Mannheim GmbH;
MWM X	70 – 12200 bp	Mannheim, Germany
100 bp DNA Ladder	100 – 1517 bp	NEB GmbH; Frankfurt am Main,
1 kb DNA Ladder	500 – 10.000 bp	Germany
Mass Ruler Ladder Low Range	80 – 1031 bp	MBI Fermentas GmbH, St. Leon-Rot, Germany
Mass Ruler DNA Ladder Mix	80 – 10.000 bp	
O' Range Ruler 100bp DNA Ladder	100 – 1500 bp	

2.1.8 Antibiotics

Stock solutions: stored at –20°C

Ampicillin (Amp 50 mg/ml): 2.5 g of ampicillin (sodium salt) were dissolved in 50 ml of 70 % ethanol.

Tetracyclin (Tet 20 mg/ml): 1 g of tetracyclin was dissolved in 50 ml of 70 % ethanol.

Kanamycin (Km 50 mg/ml): 2.5 g of kanamycin were dissolved in 50 ml of H_2O and filtered through a syringe filter (0.2 µm).

Chloramphenicol (Cm 34 mg/ml): 340 mg of chloramphenicol were dissolved in 10 ml of H_2O. Selection for a suitable resistance marker was achieved by using the following final concentrations:

Ampicillin (50 µg/ml): for culturing of BLT5403 and BLT5615 *E. coli* strains.

Ampicillin (100 µg/ml): for culturing of clones from libraries: No.988 and No.761 (*E. coli* strain).

Chloramphenicol (12.5 µg/ml): for culturing of clones from library No.958 (*E. coli* strain).

Chloramphenicol (34 µg/ml): for culturing of *E. coli* BL21 (DE3) pLysS strain.

Kanamycin (50 µg/ml): for culturing of *E. coli* BL21 (DE3) pLysS strain containing pET28a-*Taq* DNA polymerase construct.

Tetracyclin (12.5 µg/ml): for culturing of BB4 *E. coli* strain.

2.1.9 Media for growth of bacteria

All media for the culturing of bacteria were sterilized through autoclaving. Heat sensitive ingredients (glucose, maltose) were sterilized through filtration with (0.2 µm filters).

Table 2.1.9.1: Liquid and solid media used for growth of bacteria

Liquid Media	
LB-Medium (1x)	LB-Medium (2x)
Bacto tryptone 10 g	Bacto tryptone 20 g
Yeast-Extract 10 g	Yeast-Extract 10 g
NaCl 5 g	NaCl 10 g
H_2O to 1 L	H_2O to 1 L
Adjust to pH 7.0 with 5N NaOH	Adjust to pH 7.0 with 5N NaOH
SOB-Medium	SOC-Medium
Bacto tryptone 2 %	SOB-medium
Yeast-Extract 0.5 %	Glucose 20 mM
NaCl 0.05 %	
KCl 2.5 mM	
$MgCl_2$ 10 mM	
Solid Media	
TOP-Agarose / 100 ml	LB-TOP-Agar
Bacto tryptone 1 g	LB-Medium (1x) 1 L
Yeast-Extract 0.5 g	Agar-Agar 15 g
NaCl 0.5 g	
Agarose 0.6 g	

For preparation of Agar plates LB-TOP-Agar was heated in a microwave until the solution boiled and the solid agar was completely dissolved. The appropriate antibiotic(s) were supplemented after the medium cooled down up to ca. 50°C. After additional mixing, the medium was poured into the Petri dishes.

2.1.10 Buffers and solutions

Standard buffers and stock solutions that are most frequently used in this work are given in Table 2.1.10.1. Solutions that are not included in this section are given in the further sections in the context of the corresponding technique.

Table 2.1.10.1: Standard buffers and solutions used in this work

Buffers	
TBS (10x)	PBS (10x)
Tris base 30 g	NaCl 80 g
NaCl 80 g	KCl 2 g
KCl 2 g	Na_2HPO_4 x 2 H_2O 14.4 g
H_2O to 1 L	KH_2PO_4 2.4 g
	H_2O to 1 L
Adjust to pH 7.4 with HCl and autoclave	Adjust to pH 7.4 with HCl and autoclave
TBST (1x)	Tris-HCl (1 M)
TBS (1x) with 0.1 % (v/v) Tween 20	Tris base 121.1 g
	Adjust pH with HCl and autoclave
CBS (1x)	TBE (5x)
NaCl 8 g	Tris base 54 g
KCl 0.2 g	Boric acid 27.5 g
Citric acid 10.5 g	EDTA (0.5M pH8.0) 20 ml
H_2O to 1 L	H_2O to 1 L
Adjust to pH 7.0 with NaOH and autoclave	
TAE (50x)	TE (1x)
Tris base 242 g	Tris-HCl 10 mM
HAc glacial 57.1 ml	EDTA 1 mM
EDTA (0.5M pH8.0) 100 ml	
H_2O to 1 L	Adjust to pH 8.0 with HCl and autoclave

20xSSC
NaCl 175.3 g
tri-sodium-citrate 88.2 g
H_2O to 1 L
Adjust to pH 7.0 with HCl and autoclave

Buffer A
Tris-HCl pH 7.9, 50 mM
Glucose 50 mM
EDTA 1 mM

Lysis buffer
Tris-HCl 10 mM
KCl 50 mM
EDTA 1 mM
PMSF 1 mM
Tween 20 0.5 %
Nodinet P-40 0.5 %
pH 7.9

Taq storage buffer
Tris-HCl pH 7.9 50 mM
KCl 50 mM
EDTA 0.1 mM
DTT 1 mM
PMSF 0.5 mM
Glycerol 50 %

T7 phage extraction buffer (PEB)
Tris-HCl 20 mM
NaCl, 100 mM
$MgSO_4$ 6 mM
pH 8.0

λ phage extraction buffer (LD)
Tris-HCl 50 mM
NaCl 100 mM
$MgSO_4$ 8 mM
gelatine 0.01 %
pH 7.5

SM PCR buffer (10x)
KCl 100 mM
$(NH_4)_2SO_4$ 60 mM
Tris-HCl pH 8.8 200 mM
$MgSO_4$ 20 mM
Triton X-100 1 %
BSA 1 mg/ml

Membrane Blocking Solution (MBS)
20 ml Casein Based Blocking Buffer Concentrate, (Genosys Biotechnologies, Cambrige, UK)
TBST pH 8.0 80 ml
Saccharose 5 g
The resulting pH will be 7.0, store at 4°C

Solutions

EDTA 0,5 M stock solution
EDTA 186.1 g
H_2O to 1 L
Adjust to pH 8.0 with NaOH and autoclave

IPTG 1M stock solution
IPTG 2.38 g
H_2O to 1 L
Sterilize by filtration (0.2 µM), store at –20°C

BCIP (stock solution)
BCIP 60 mg
DMF 1 ml
Store at –20°C, protect from light

MTT (stock solution)
MTT 250 mg
DMF 70% in H_2O 5 ml
Store at –20°C, protect from light

BCIP/MTT substrate solution for AP	Sodium azide stock solution (2.5 %)
CBS 20 ml	NaN_3 2.5 g
$MgCl_2$ (1 M) 100 µl	H_2O 100 ml
BCIP 40 µl	Protected against light
MTT 60 µl	
Sodium acetate 3 M stock solution	Betaine 5 M stock solution
NaAc x 3 H_2O 408.1 g	Betaine 58.55 g
H_2O to 1 L	H_2O to 100 ml
Adjust to pH 5.2 with glacial HAc and autoclave	
Alkaline lysis solution I	Alkaline lysis solution II
Glucose 50 mM	NaOH 0.2 M
Tris HCl (pH 8.0) 25 mM	SDS 1 %
EDTA (pH 8.0) 10 mM	
Autoclave and store at 4°C	
Alkaline lysis solution III	Ortho-phenylenediamine solution (OPD)
Potassium acetate 3 M	OPD 1 mg/ml
HAc glacial 5 M	H_2O_2 10 µl/ml (0.03 %)
Solutions for SPOT synthesis	
Bromophenol blue (BPB) stock solution	Acetylation mix
10 mg BPB / 1 ml DMF	2 % solution of acetic anhydride (p.a) in DMF
Piperidine (p.a.)	Deprotection mix
20 % in DMF	50 % TFA, 3 % TIBS, 2 % water, 45 % DCM

2.1.11 Fmoc-amino acid stock solutions

Fmoc-amino acid derivatives of all 20 L-amino acids as well as ß-alanine and other special amino acid derivatives are available from several suppliers in sufficient quality (Novabiochem/Merck Biosciences, or Bachem). Side chain protecting groups should be Cys(Acm) or Cys(Trt), Asp(OtBu), Glu(OtBu), His(Trt), Lys(Boc), Asn(Trt), Gln(Trt), Arg(Pmc), Ser(tBu), Thr(tBu), Trp(Boc), and Tyr(tBu). HOBt-esters of these amino acid derivatives must be prepared in NMP for use throughout in spotting reactions (see 2.2.6 III).

2.2 Methods

2.2.1 Growth of bacteria

All bacterial strains were grown in LB-medium at 37°C under air agitation conditions with supplement of appropriate antibiotics (see 2.1.8). For this purpose, different lab-ware was used. Small overnight cultures that originate from a single colony were performed in 10 ml Capsenberg tubes containing 5 ml LB-medium and cultured in an incubator with rotation (200 rpm). Larger cultures were grown in 0.1 ml to 1.0 L baffled flasks containing LB-medium with rotation (150 rpm) until required OD_{600} = 0.5 or 1.0 was reached. The volume of bacterial medium didn't exceed the 1/5 of the total flask volume. Bacteria from clone libraries were grown in 96 deep well blocks (Qiagen), at 37°C, overnight with high rotation speed (300 rpm). For high-throughput growth, the cultures were initiated by inoculation of 1.6 ml of 2 x LB medium (in 96 well blocks) with a drop of the bacterial clone directly from a glycerol stock. For this purpose the 96-pin plastic replication tools (BioCat) were used. The blocks were covered with adhesive gas permeable tape (AB Gene) and incubated under the above conditions. The *E. coli* BB4 (plating bacteria) that is a host for the lambda phages were grown up to OD_{600} = 1.0 with a supplement of maltose and $MgSO_4$ to the final concentrations 0.2 % and 10 mM respectively. Bacteria were harvested by centrifugation at 4000 x g and the resulting pellet was redisolved in 50 % of the original culture volume of 10 mM $MgSO_4$. Host strains for T7select10-3b phage were grown as follows. The BLT5615 were grown until OD_{600} = 0.4-0.5 was reached, then supplemented with IPTG (final concentration 1 mM) and grown for about 30 minutes (OD_{600} = 1). The BLT5403 bacteria were grown without IPTG supplement, until OD_{600} = 1.0. The additional information considering culturing that is not given here will be specified in the corresponding sections.

2.2.2 Storage of bacteria and phages

Bacterial colonies were obtained by streaking out onto agar plates using an inoculation loop and kept at 4°C for less than one month. Liquid cultures were stored at 4°C for no longer than one week. For long time storage, fresh log phase cultures were mixed with sterile glycerol (final concentration 20 %) in sterile 1.5 ml cryo-resistant tubes and kept at –70°C. Replicas of library clones were stored in 96 well plates covered with cryo-resistant adhesive foil. Clarified phage lysates were stored at 4°C approximately up to two months. For long-term storage or as a backup, samples of the lysates were combined with 0.1 volume of sterile 80 % glycerol and then frozen at –70°C.

2.2.3 Molecular biology methods

2.2.3.1 Precipitation of DNA

DNA precipitation was mediated by high salt concentration and addition of either isopropanol or ethanol. Precipitation with ethanol was carried out by addition of sodium acetate up to the final concentration 0.3 M (pH 5.2) followed by addition of 2 volumes of the cold 96 % ethanol and thorough mixing. After 30 min incubation on ice, the DNA was recovered by centrifugation at 15000 x g, 4°C for 30 min. With the supernatant removed, the pellet was washed with 70 % ethanol (-20°C) and the centrifugation step was repeated for 15 min. The ethanol was removed and the pellet was dried either with air on a sterile bench or using vacuum evaporator with rotation but without heating. The DNA was redissolved in an appropriate volume of Milli-Q water or TE buffer. For probes where an increase in volume was undesirable, precipitation with isopropanol was carried out. The ammonium acetate was added to the sample up to the 2.5 M final concentration followed by the addition of 0.7 volumes of isopropanol. The content was mixed appropriately and incubated at room temperature for 30 min. All conditions and centrifugation steps were the same as described above for the ethanol procedure. Alternatively, only precipitation with isopropanol was carried out. For recovery of plasmid DNA (see 2.2.3.6) and PCR products amplified from human brain amygdala clone library, 0.7 – 1.0 volumes of isopropanol were precisely mixed with each sample (in 96 well PCR plates or deep well blocks). Then they were incubated for 45 min – 1 h at room temperature and centrifuged at 3500 x g for 45 min, at room temperature. The supernatant was discarded and residual drops of a liquid removed by tapping on a paper towel. The DNA was air dried for 30 min, at room temperature and redissolved in the appropriate solution.

2.2.3.2 Determination of DNA concentration

DNA concentration was measured using GeneQuant II RNA/DNA Calculator (Pharmacia Biotech) in the 10 µl Hellma quartz suprasil glass cuvette (10 mm). Preceding usage, the cuvette was washed several times with Milli-Q water. Milli-Q or appropriate buffer was used for the calibration of the photometer. If necessary, samples were diluted and the dilution factor was listed in the GeneQuant record. The DNA concentration was obtained by multiplying the absorbance (A260 nm) by the factor of dilution, and by the nucleic acid conversion factor:

DNA Concentration [µg/ml] = A (260 nm) x dilution factor x F
F = 50 (ds DNA), 33 (ss DNA), 40 (RNA).
Alternatively, the approximate DNA concentration was determined via AGE analysis, by comparison of the fluorescence of sample band with the fluorescence of bands of the correct molecular mass marker that was run on the same gel.

2.2.3.3 DNA agarose electrophoresis

Agarose gels (0.6 – 2 %) were prepared by dissolving 0.6 – 2 g of peq-GOLD universal agarose in 100 ml of 1 x TBA or TAE buffer by boiling the mixture at least three times in a microwave. For 2 – 3 % preparative gels, the appropriate amounts of the low melting point agarose were used. The agarose solution was allowed to cool down to approximately 60°C and 5 µl of ethidium bromide solution (10 mg/ml) were added. Agarose was poured into appropriate gel chambers containing a comb for the formation of sample pockets and left at room temperature for polymerisation. Ready to use gel was transferred into the electrophoresis device and covered with a running buffer. Always, the same buffer that was used for gel preparation was also used as a running buffer.
Depending on the comb used for the pockets formation, different volumes of probe were applied, ranging from 5 µl to 400 µl. Before application onto the gel, DNA samples were mixed with 0.2 vol of 6 x sample loading buffer (40 % sucrose in water, 0.25 % Bromophenol-blue, 0.25 % of xylene-cyanol FF, optional). Additionally a suitable DNA marker was loaded. DNA fragments were separated at 60 – 90 V for long-term runs and 120 – 170 V for short-term runs. For long-term electrophoresis the TAE buffer was used preferentially. The different bands were detected on a transilluminator (302 nm). The gels were documented with the Herolab CCD camera connected to a PC and controlled by EASY image 3.16. The images were saved as 8 bit TIFF files.

2.2.3.4 DNA extraction from agarose gels

For isolation of DNA of a particular size from a DNA mixture, the different DNA fragments were separated on low melting point agarose gels. The piece of the agarose gel containing the desired band was cut out on a transilluminator with a scalpel. The DNA was then extracted from the gel with the Qiaquick Gel Extraction kit (Qiagen) according to the manufacturer’s guidelines.

2.2.3.5 Isolation of phage DNA

The λDisplay1 phage vector DNA was isolated from the phage lysate using the Nucleobond AX-100 kit (Machery-Nagel) according to the manufacturer's protocol. The DNA of single clones or phage DNA mixtures from T7 human brain library was isolated from phage lysates (see 2.2.4.2). This preparation was used for isolation of phage DNA that served as a template for PCR reactions. The volume of 50 – 200 µl of phage lysate were transferred into sterile Eppendorf tubes and mixed with EDTA (final concentration 10 mM) and then placed in a thermo mixer at 65°C for 10 min with moderate shaking. The mixture was cooled down to room temperature for 3 – 5 min and centrifuged at 14000 x g for 3 min. Either the aliquot of 2 µl of prepared phage DNA solution was directly used for 100 µl "check-PCR" reaction, or the whole supernatant was filtrated through the MultiScreen filter plate (Millipore) and washed with sterile water twice, in order to eliminate the EDTA from the mixture. The DNA was taken up from the filter by redissolving it in 65 – 100 µl of sterile Milli-Q water.

2.2.3.6 Isolation of PCR products and plasmid DNA

PCR products were purified and isolated with the QIAquick PCR purification (Qiagen) or using the MultiScreen PCR Plates (Millipore) according to the producer's manual.

Plasmid DNA was isolated with differing methods, which are all based on the alkaline lysis method (Birnboim & Doly, 1979). For the purpose of this work only small-scale preparations were performed (up to 20 µg). Ion exchange columns were commonly used for plasmid preparation (QIAprep Spin Miniprep Kit, Qiagen). For a larger number of probes two commercially available kits were used, R.E.A.L Prep 96 plasmid kit (Qiagen) and Montage 96 plasmid miniprep kit (Millipore). Plasmids from 6144 clones of the human brain amygdala cDNA library (RZPD No 761) were isolated using a shortened alkaline lysis protocol. All solutions used for this preparation are described in Table 2.1.10.1 Library clones were grown in 96 deep well blocks, as described in 2.2.1. Bacteria were harvested by centrifugation at 4°C, 3000 x g for 15 min. The supernatant was decanted from the culturing block. The residual medium was removed by tapping the block on the paper towel. Using a multi-channel pipette, 100 µl of alkaline lysis solution I were added. Wells of the block were covered with a tight rubber cover and the cell pellet was precisely re-suspended by vortexing. The block was kept for 5 min at room temperature. Next, 200 µl of solution II were added and mixed very gently. The block was immediately placed on ice for 5 min. After addition of 150 µl of solution III and intensive shaking, the block was again placed on ice for 5 min, followed by centrifugation (as above). The clear supernatant (~ 400 µl) was transferred into the block containing 400 µl of isopropanol, mixed exactly with the pipette. Plasmid DNA was

precipitated with isopropanol only, as described in 2.2.3.1 and redissolved in 100 µl of Milli-Q water.

2.2.3.7 DNA digestion

DNA was digested with restriction enzymes from New England Biolabs, using the recommended buffer. For double digestion the optimal buffer was chosen according to the manufacturer's recommendations. As an option, the digestion was run separately for each enzyme with a purification step in between. Amounts of endonucleases used for DNA digestion are given in units ("U" according to producer's definition). Usually 1 U of enzyme was used for each µg of DNA. The DNA solution was mixed with 0.1 vol of appropriate digestion buffer (10x) and appropriate amount of endonucleases/s. The mixture was incubated at 37°C for 1 – 4 h. The total reaction volume was dependent on the amount of enzyme/s used for digestion and exceeded 10 times the total volume of enzyme/s. This restricted volume was, increased when the sensitive endonucleases were used (e.g. *Hind*III) in order to avoid their star activity. Overnight incubations were applied for digestion of PCR products that contained the cleavage sites in a close proximity to the strand termini. The amount of required endonucleases was increased as well (10 – 20 U). Usually enzymes were inactivated by 20 min incubation at 65°C. According to the future applications and fragment size, the digested DNA was purified either by L.M.P gel electrophoresis followed by extraction using Qiaquick Gel Extraction kit (Qiagen), using QIAquick spin columns (Qiagen) or by alcohol precipitation. To digest the PCR amplified and Cy-dye labelled inserts from the human brain phage library that were used as targets for microarray hybridization, 5 U of the *EcoR*I and *Hind*III enzymes were added directly to the PCR reaction. The mixture was supplemented with the 12 µl of the buffer for *Eco*RV or buffer 2 (NEB) and sterile Milli-Q water up to 120 µl. The reaction was run at 37°C, overnight.

2.2.3.8 Dephosphorylation

Digested DNA fragments that were used for ligation were treated with alkaline phosphatase from shrimps (USB cooperation, Cleveland, USA). The alkaline phosphatase removed 5' phosphate groups from the vector or insert and therefore prevented so called self-ligation of inserts or vector closure. 0.1 volumes of the alkaline phosphatase buffer were mixed with the DNA sample followed by the addition of enzyme (1 U for each pmol of DNA). The mixture was incubated for 1 h at 37°C. The alkaline phosphatase was inactivated for 15 min at 65°C. The mixture was ready to use for ligation reactions.

2.2.3.9 Ligation

The T4 DNA ligase (Promega) was used for cloning of the DNA sequences coding the model protein domains (EVH1, WW/FE65, WW/YAP) into the plasmid vectors. The plasmid vectors were previously digested and purified from the L.M.P agarose gel. The appropriate inserts were PCR amplified, digested, dephosphorylated and purified from the L.M.P agarose gels. The molar ratio of insert to vector used for ligation was 3:1. The amounts of 0.8 – 1.5 µg of total DNA mixture were mixed with the 0.1 volumes of appropriate buffer and supplemented with 5 U of the T4 DNA ligase. The ligation reaction was carried out at 16°C, overnight in 10 µl volume. 10 µl of water were added to the reaction and heated up to 65°C for 15 min for increased electroporation efficiency (Ymer, 1991). 20 ng of the ligated DNA was taken for transformation. Ligation of DNA coding for the model domains with the λDisplay1 phage vector was performed as follows. Previously PCR amplified, digested with *Spe*I and *Not*I endonucleases and purified DNA fragments were ligated with the λDisplay1 vector arms. The vector arms were previously separately digested with *Spe*I and *Not*I, dephosphorylated and purified by precipitation (for cloning strategy refer to 3.1.1). A 15 µl ligation reaction contained: 6 U units of T4 DNA ligase, 25 µg/ml acetylated BSA, 10 mM DTT, 1 mM ATP, 4.2 µg of λDisplay1 vector arms and 11.6 ng of DNA from each domain insert. The ligation was carried out overnight at 16°C. 8 µl of the ligation mixture was directly used for phage assembly using commercially available “Ready To Go” kit (see Table 2.1.1.1), according to the manufacturer’s protocol.

2.2.3.10 Transformation of *E. coli* cells

The transformation of the Gene Hogs *E. coli* cells was performed through electroporation, using the Bio-Rad Gene Pulser. 50 µl of the electro competent cells were carefully mixed (by agitation) with a ligation mixture (approximately 20 ng of DNA) and incubated on ice for 2 min. This mixture was transferred into the ice-cold 2 mm electroporation cuvette (Bio-Rad) and electroporated with the following parameters: 2.5 kV, 25 µF and 200 Ω, as described by Dower *et al*., 1988. The time constant was 4.5 s. After electroporation the cells were immediately mixed with the 200 µl of prewarmed SOC medium (37°C), transferred to the 5 ml culturing glass tube and grown at 37°C for 1 h. The aliquots of this culture (10 µl, 100 µl, 1000 µl) were plated out onto the LB agar plates containing appropriate antibiotics and incubated at 37°C overnight.

2.2.3.11 Polymerase chain reaction

PCR was used for amplification of genes flanked by the appropriate restriction sites, for analysis of phage clones, for amplification of DNA fragments from clone libraries that were used for DNA microarray fabrication and amplification and labeling of phage target DNA used for hybridization with the microarray chip. Several different DNA polymerases were used including hot start, *Taq* and *pfu* as well as the self-made *Taq* DNA polymerase, expressed and purified in the lab (see Table 2.1.1.1) PCR reactions were carried out in a MJ research PTC 200 Thermocycler using 0.2 ml strips (AB Gene; Epsom, Surrey, UK) or in a MJ tetrad Thermocycler with 4 alpha blocks using 96 well plates (AB Gene). Several different protocols were developed for different proposes. All are based on the protocol of Sambrook *et al*., (1989). The most commonly used protocols are described below. The modifications that are not described here are given in the context of a corresponding experiment. The different primers used for different PCRs are listed in Table 2.1.6.1.

The PCR mixture for amplification of the DNA fragments from the T7Select10-3b phage vector. Different polymerases were used, according to availability (Table 2.1.1.1).

PCR reaction per 100 µl:		**final concentration**
DNA-polymerase buffer (10x)	10 µl	1x
$MgCl_2$ (25 mM)	6 µl	1.5 mM
dNTP mix (10 mM)	2 µl	0.2 mM (each)
Sense primer TS 21 (10 pmol)	2 µl	20 pmol
Antisense primer TA 20 (10 pmol)	2 µl	20 pmol
Taq-DNA-polymerase	0.5 µl	2.5 U
H_2O to 100 µl	dependent on the template volume	
DNA template 2-50 µl	dependent on the experiment	

The PCR program:

1. initial denaturation:	94°C for 2 min
2. denaturation:	94°C for 30 s
3. annealing:	49°C for 30 s
4. elongation:	72°C for 3 min
2-4:	35 cycles
5. final extension:	72°C for 6 min
6. end	4°C

In the case of HotGoldstar DNA Polymerase usage, the initial activation step was applied according to the manufacturer's recommendations (5 min at 95°C).

The PCR amplification and labeling of the DNA fragments from T7 human brain library was carried out using the program described above. The PCR mixture was slightly changed. Self-made *Taq* DNA polymerase was used in combination with the *pfu* DNA polymerase. Both enzymes were mixed in an average ratio of 10 : 1 units respectively. The buffer used for PCR was made according to Lu & Erickson, 1997 (SM PCR buffer: 100 mM KCl, 60 mM $(NH_4)_2SO_4$, 200 mM Tris-HCl, pH 8.8, 1 % Triton X, 1 mg/ml BSA).
Final concentrations of the deoxynucleotides: dATP, dGTP, dTTP were 0.2 mM. The dCTP was added to the final concentration of 0.19 mM. The PCR mixture was supplemented with 1 µl of Cy3-dCTP or Cy 5-dCTP (final concentration 0.01 mM).

The PCR mixture for amplification of DNA fragments from the full-length cDNA human brain amygdala library:

PCR reaction per 100 µl:		**final concentration**
SM PCR buffer (10x)	10 µl	1x
$MgSO_4$ (25 mM)	16 µl	4.0 mM
dNTP MIX (10 mM)	4 µl	0.4 mM (each)
Sense primer M13 for 23 (10 pmol)	2 µl	20 pmol
Antisense primer M13 rev 30 (10 pmol)	2 µl	20 pmol
Betaine (5M)	20 µl	1 M
SM *Taq*-DNA-polymerase	0.3 µl	1.5 U
pfu DNA polymerase 2.5 U/µl	0.08 µl	0.15 U
H_2O	45 µl	
Template (plasmid DNA)	1 µl	50 – 100 ng

The PCR program:

1. initial denaturation:	94°C for 2 min
2. denaturation:	94°C for 30 s
3. annealing:	56°C for 30 s
4. elongation:	72°C for 2.3 min
2-4:	10 cycles
5. denaturation:	94°C for 33 s,
6. annealing:	56°C for 33 s

7. elongation:	72°C for 2.33 min
5-7:	20 cycles
8. final extension:	72°C for 5 min
9. end	4°C

2.2.4 Phage work

2.2.4.1 Preparation of λDisplay1 phage lysates

The original λDisplay1 phage lysate was diluted in the LD buffer (see Table 2.1.10.1) in order to obtain the appropriate phage concentration of 10^4 pfu/ml. 400 µl of diluted phages were mixed with 400 µl of *E. coli* BB4 plating bacteria (OD_{600} = 2.0 in 10 mM MgSO4) and incubated for 20 min at 37°C. Then 12 ml of LB-top agarose was added, the mixture was poured onto 145/20-mm-large Petri dishes and incubated at 37°C overnight. Phages were eluted from the agarose by filling up the dish with 20 ml of SM LD buffer followed by incubation for 5 h at 4°C, under gentle agitation on a rocker table. The liquid was decanted, supplemented with chloroform (2 % final concentration) and centrifuged at 4°C with a speed of 4500 x g. The phage lysate was separated from the pellet and poured into a new tube.

2.2.4.2 Preparation of T7select10-3 phage lysates

Phages displaying proteins from the normal human brain cDNA library were cultured using the BLT5615 *E. coli* host, whereas phages displaying model protein domains were grown using the BLT5403 host (see 2.2.1). An aliquot from the original phage lysate (1:10^3 volume of a future culture) was added to the ready for infection host culture and grown at 37°C with agitation (200 rpm), until lysis occurred. An increase of the culture transparency and an emergence of the bacterial debris indicated the complete lysis. The lysates were clarified by spinning at 10000 x g and 4°C and decanted into sterile tubes. Additionally, the individual phages were extracted from a single plaque that was placed in an eppendorf tube with 50 – 200 µl of the phage extraction buffer (PEB) and incubated for at least 2 h at 4°C.

2.2.4.3 Determination of the T7 and lambda bacteriophage titers

The phage titer, described in plaque forming units per unit volume (e.g. pfu/ml) is the number of phage plaques visible on the plate multiplied by an appropriate dilution and multiplied by an aliquot of the volume (e.g.10, to account for the 0.1 ml of dilution plated). For example, if the aliquot of 100 µl from the phage dilution 1:10^6 was poured onto the agar plate and 30 phage plaques were obtained, then the titer must be calculated according to the following

formula: 30 x 10^6 x 10 = 3 x 10^8 pfu/ml. To calculate a number of plaque forming units (pfu) in 1 ml of lysate, a series of dilutions of an appropriate sample were prepared using LB medium or LD buffer as the diluent for the T7 phage and the λ phages respectively. The initial 1:100 dilution was prepared by adding 10 µl of sample to 990 µl of medium (1:10^2 dilution). Serial dilutions were made by adding 100 µl of the 1:10^2 dilution to 900 µl medium (1:10^3 dilution), 100 µl of the 10^3 dilution to 900 µl medium (10^4 dilution), and so on. A series of 10 ml sterile tubes were prepared by pipetting 250 µl and 100 µl of the host cells for T7 and λ phage respectively. Starting from the highest dilution, 100 µl of the phage dilution were mixed with the bacteria. 2.5 ml of molten Top-agarose (45°C) were added to each probe, mixed well and poured immediately onto a pre-warmed (37°C) agar plate. This step was preceded by addition of IPTG (final concentration 0.5 mM) to the mixture, in case of the T7 phages. In case of the λ phage addition of Top-agarose was preceded by incubation of phage and bacteria mixture for 20 min at 37°C. After the contents were poured out, plates were allowed to sit undisturbed for several minutes until the top agarose hardened, then inverted and incubated for 3–4 h at 37°C or overnight at room temperature (T7 phage) or overnight at 37°C (λ phage). The phage plaques were counted and the phage titer was calculated as described above.

2.2.4.4 Affinity selection of phages on the SPOT membrane

A piece of the SPOT cellulose membrane containing 4 peptides with different sequences was placed in a 2 ml Eppendorf tube and washed for 1 h with the 70 % ethanol. Ethanol was decanted and the membrane was washed 20 times for 5 min with TBS pH 7.4, at room temperature under agitation on a rocker table. Then the membrane piece was blocked for at least 1 h with 2 ml of blocking solution [5 % skim milk, 50 µg/ml fish sperm DNA, 50 µg/ml heparin in 0.1 % TBST pH 7.4, 0.1 % NaN_3]. 0.9 ml of phage lysate supplemented with protease inhibitor were blocked with 0.1 ml of 5 x concentrated blocking solution for 10 min at room temperature under gentle agitation. The membrane was placed in a 2 ml tube and incubated with blocked phages overnight at 4°C under gentle agitation. The extensive washing steps were performed to eliminate the unspecific or unbound background phages. The membrane was washed 10 times for 5 min, with 0.5 x concentrated blocking buffer and 10 times for 5 min with 1 x TBST pH 7.4 (see Table 2.1.10.1). The individual spots were cut out and transferred to 0.5 ml tubes containing 100 µl or 150 µl of the elution reagent, in case of the T7 phages 1 % SDS or 1 M guanidine. Elution for the T7 phages was carried out at room temperature for 10 min with gentle shaking. Affinity selection of the λ phages on the peptide-spot membrane was done in a similar manner, except for the elution step. The λ phages were eluted via direct infection of 150 µl of the BB4 host bacteria at 37°C for 20 min.

Eluted phages were used for lysate preparation, phage tittering or DNA isolation. T7 human brain phage displayed library was panned on the cellulose membrane containing 425 peptides-spots as described above. Nevertheless a certain alterations regarding the volumes of the phage lysate as well as the blocking and washing buffers were introduced. The whole membrane was placed in a plastic box, washed with 30 ml of ethanol and TBS as described above and blocked with 30 ml of the blocking solution for 1 h at room temperature with agitation. Following blocking, 30 ml of the phage displayed library was panned on the spot membrane under the same conditions as described above. Washing of the membrane was done as above using 30 ml of each solution for each washing step. Finally, the membrane was placed in a square dish filled up with the TBST buffer. The peptide spots were cut out with a scalpel and phages were eluted as follows. The first row with 25 peptides was separated from the whole membrane and placed onto a square plate filled with the TBST buffer. The individual spots were cut out one by one and placed in a 96 well plate filled with 75 µl of TBST. When all 400 spots were separated and placed in the 96 well plates, 75 µl of 2 % SDS was added and the phages were eluted simultaneously at room temperature for 10 min, with a gentle shaking. 150 µl of phage eluates were transferred into 96 deep well blocks containing 1350 µl of the LB medium. A 10 µl aliquot was taken from each well for preparation of the appropriate dilution ($1:10^4$). Then 100 µl of this last dilution was poured onto the standard Petri agar plates as described in 2.2.4.3.

2.2.4.5 Affinity selection of phages on peptide synthesized on a plastic foil surface

Control phage panning on peptides synthesized on the new type of plastic foil surface was prepared in 2 ml Eppendorf tube as described above for the cellulose spots. T7 phages displaying model protein domains (EVH1, WW/FE65, WW/YAP) were diluted in the mixture of T7 phages displaying IL4 protein (see section 3.6). The pieces of foil with peptide patches as well as the phage mixture were blocked as described for the cellulose spots. The number of washing steps during the panning procedure was reduced to 5 times for 5 min for both solutions. The elution step, for peptide-patches that were cut out from the foil, was carried out using 100 µl of 1 M guanidine solution, as described above. The phage eluates were ready for dilution preparation and titering.

2.2.4.6 Target Assisted Iterative Screening (TAIS)

The human brain T7 phage display cDNA library was panned on the cellulose peptide-spots according to the protocol given in 2.2.4.4. Phages were eluted from the peptide-spots with 100 µl of 1 M guanidine. $1:10^3$ and $1:10^4$ dilutions of post-panning phage populations in LB

medium were performed in order to achieve separated phage plaques after plating onto 145/20-mm-large agar plates (see 2.2.4.3). The dilution is highly dependent on the titer of the phage sample eluted from the peptide-spot and must be experimentally established for each experiment. For plating, 100 µl of diluted phage eluate was mixed with 750 µl of BLT5615 host bacteria (ready for plating) and 9 ml of top-agarose followed by the addition of IPTG (final concentration 0.5 mM). The plates were incubated for 3 – 4 h at 37°C. Probing of the pre-selected library subset on a membrane was done according to a protocol developed by Kurakin, (Kurakin and Bredesen, 2002), that was slightly modified. To prevent top-agarose from sticking to the nitrocellulose membrane, plates were cooled down at 4°C for 30 min. The nitrocellulose "protran" membrane (132 mm ϕ) (see Table 2.1.1.1) was placed onto the plate with a special care taken to avoid trapping air bubbles and left for 5 min at room temperature. Three deep holes (coordinate points) were made in two places of the membrane perimeter using a red-hot syringe needle. The membrane was placed in a new large Petri dish and blocked with 1 % BSA for 1 h at room temperature with gentle rocking. In the meantime, the biotinylated peptides were mixed with a streptavidine-alkaline phosphatase conjugate (STRAP) at the molar ratio of 4:1, in 100 µl of TBST pH 7.4 and incubated for 15 min at room temperature. The target-STRAP conjugate was diluted in 30 ml of TBST containing 0.5 % BSA. The blocked membrane was then incubated with this mixture overnight at 4°C with gentle rocking. Subsequently, the membrane was transferred into a fresh dish and washed 3 times for 15 min with 50 ml of TBST, at room temperature on a rocker. After these washing steps, 20 ml of an alkaline phosphatase substrate solution BCIP/MTT (see Table 2.1.10.1) was freshly prepared and the membrane was incubated at 4°C without shaking in the dark, until plaques became visible (up to 3 h). The staining solution was decanted and the membrane was washed a few times with TBST and stored at 4°C in the same buffer. With a help of the coordination points, the phage plaques on the agar plate, corresponding to the positive phage plaques on the membrane (blue signals), were picked up with a sterile tooth stick and transferred into a sterile Eppendorf tube containing 200 µl of PEB extraction solution. Phages were extracted at 4°C for 2 h (see 2.2.4.2). A 100 µl of the extracted phage solution were taken for the DNA isolation purpose (see 2.2.3.5) that was used then as a template for PCR. Another 100 µl was used for phage propagation (see 2.2.4.2).

2.2.5 Fabrication and handling of DNA microarray biochips

The nomenclature used in this work, for the description of the nucleic acids that are either in the solution or tethered on the solid support is according to (Phimister, 1999). The nucleic acid that is spotted on the glass slide is described as a "probe" and the labeled free nucleic acid as a "target".

2.2.5.1 Preparation of DNA samples for printing

After PCR amplification (see 2.2.3.11) and isopropanol precipitation (see 2.2.3.1), 6144 genes (amygdala library) to be spotted on the glass slides were recovered by addition of 12 µl of spotting buffer (3 x SSC, 1.5 M betaine). The plates were centrifuged by a short spin and left at 4°C overnight. 10 µl of each DNA solution were transferred from 96 well plates into 384 well source plates (AB Gene), using a multi channel pipette and the 10 µl, thin pipette tips (Labsystems). The source plates were centrifuged by a short spin and stored at 4°C for use.

2.2.5.2 Printing of high density DNA microarray

PCR fragments were printed on amine silanated slides (Cell Associates Inc) and super amine slides (TeleChem). For array printing the contact microarrayer "MicroGrid II" (BioRobotics) and 16 split pins "MicroSpot 2500" (GeneWorx) were used. The split pins "resting" in a printing tool were placed in a small flask with sterile water and cleaned by sonication for 30 min. They were then transferred to a second flask containing 70 % p.a ethanol. Before DNA spotting was initiated, the pins were primed by test spotting with a buffer solution, where each pin struck the glass slide surface over 4000 times. Printing conditions were established experimentally and are shown in the appendix 1. The arrays were left in the microarrayer or at 22°C for a few hours after printing completion. A label on each microarray slide was inscribed with a diamond pen before a slide processing started.

2.2.5.3 Fixation of the DNA features to the slide surface

The DNA features were re-hydrated by lightly moistening the array face of each slide in the vapour of a warm (37°C) water bath for about 15 – 20 s (Bowtel *et. al* 2003). The slides were snap-dried immediately by heating for 1 – 3 s on a clean hot plate (120°C). The DNA was fixed to the slide substrate by two times UV cross-linking at 1200 µjoules using the Stratalinker 1800 device (Stratagene). The slides were then subjected for a chemical blocking or blocking with a BSA solution, as described below. The chemical blocking was carried out immediately after the DNA fixation procedure and the ready to use slides were stored in a dry, clean plastic box. Blocking with the BSA solution was performed always immediately before a microarrays application. Until that time, the biochips were stored under the same conditions as described above.

2.2.5.4 Washing and blocking the slides with BSA solution

All biochips that were blocked with this method had to be used for hybridization within the time period of 30 minutes after blocking to avoid BSA denaturation on their surface. The slides were placed in a slide rack and immersed into a glass jar containing clean, 95°C hot water for 2 min and then washed several times with room temperature water, in a similar manner. Free amino groups on the slide surface were blocked by incubation of the slides for 45 min at 42°C in a prewarmed solution containing filtered BSA solution (0.5 % BSA, 5 x SSC, 0.1 % SDS). To remove unbound DNA and blocking components (especially SDS), the microarray substrates were washed several times in water. The slides were dried with a stream of N_2 (pressure 1.2 bar), by holding the slide with a forceps and streaming the gas in the opposite direction. This process was carried out quickly for both sides of the slide. A slow drying process would result in a high background (smear appearing on the glass slides). The prepared biochips were used within 15 min for hybridization.

2.2.5.5 Chemical blocking with succinic anhydride

The blocking solution was freshly prepared by dissolving succinic anhydride in 200 ml of anhydrous DCE followed by addition of 2.5 ml of *N*-methylimidazole (Diehl *et.al*. 2001). The slides were blocked for 1 h at room temperature with gentle rocking and washed briefly in 200 µl of fresh DCE. Immersing the slides several times in 95 % ethanol washed off the DCE. For DNA denaturation, the slides were kept for 2 min in 95°C water, as described before. The slides were immediately washed, first in water (10x) then in ethanol (10x) followed once again by wash in water (10x) and subsequently dried with a stream of N_2 or by centrifugation either in the 50 ml falcon tubes or in the centrifugation racks for 5 min at 1100 rpm. Care was taken not to allow the slides to dry out during the washing steps. The slides were stored as described above.

2.2.5.6 Cleaning of cover-slips

The 24 x 50 mm cover-slips used for hybridization (IDL Biotech AB, Bromma, Sweden) were immersed in isopropanol and water several times and cleaned with a special wipe (provided with the Telechem slides) that didn't leave any contaminants. After the last isopropanol wash, the cover-slips were dried with a stream of N_2. The whole procedure was repeated if any contamination was still visible.

2.2.5.7 Array quality check hybridization

To check the quality of the spotted array, hybridization with the Cy3 labeled random primers (15mere) was performed. A 3 µl of random primer mixture was mixed with 27 µl of the appropriate hybridization buffer (1 x SSC, 0.1 % SDS). The DNA solution was dispensed near one of the long ends of the microarray slide. The cover-slip (IDE) was held with forceps on the one side and the edge of the second side was placed just behind the drop of the DNA solution and held with a second pair of forceps to prevent moving. The cover-slip was lowered until the liquid covered the total arrayed area. Care was taken to avoid air bubble formation. The biochip was placed in a humidity chamber. A few µl of hybridization buffer were deposited in each corner of the chamber to keep the moisture level constant. The biochip was tightly locked in the hybridization chamber and the hybridization was carried out at room temperature overnight.

2.2.5.8 Target to probe hybridization

The pre and post panning phage DNA populations were PCR amplified and labeled as described above. The DNA concentration for each target was determined as described in 2.2.3.2. When a two-dye system was used, the same quantity (ng) of post panning (Cy3) and pre panning (Cy5) targets were mixed together and supplemented with 400 ng of the control DNA (Fas), also labeled with Cy5 and Cy3 and mixed in a 1:1 ratio. This DNA mixture was applied on a microarray biochip for hybridization. In the case of a one-dye system, both targets were labeled with the Cy5 dye and the same quantities (ng) of the pre and post panning samples were mixed separately with the Cy5 labeled control Fas DNA (400 ng). Two separate targets were applied onto the two biochips for hybridization. The quantities of single targets to be hybridized with a biochip ranged from 2 to 4 µg. The target DNA mixtures were supplemented with water (if necessary) and 1 vol of 2x concentrated hybridization buffer (5 x SSC, 0.2 % SDS). The total volume of sample that was ready for hybridization was 200 µl. The hybridization was carried out in the Lucidea Slide Pro hybridization station (Amersham). Before each hybridization run, the chambers and pipes of the device were cleaned twice with hot Milli-Q water using a wash protocol designed for cleaning. A 180 µl out of 200 µl of the hybridization mixture was injected into the hybridization chamber containing a microarray biochip inside, using a Hamilton syringe. Hybridization was carried out for 16 hours at 42°C with mixing every 15 min (mixing: volume 30 µl, speed 10 µl/s). Pipes and chambers were cleaned with twice with 500 µl of water (speed 150 µl/s) followed by cleaning with isopropanol under the same conditions (x4). Microarrays were washed as described below. For a manual hybridization using cover-slip, the mixture of target DNA was dried out using a evaporator (without heating) and redissolved in 30 µl of the hybridization buffer (2.5 x SSC, 0.1 % SDS),

overnight at 4°C. The target DNA solution was deposited on the microarray biochip as described in 2.2.5.7 and the hybridization was carried out at 42°C for 16 h. Before each hybridization procedure, the DNA was additionally denaturized by heating at 95°C for 2 min and then cooled down to room temperature.

2.2.5.9 Post hybridization slide processing

Microarray slides were processed immediately after the hybridization procedure was completed. Special care was taken to avoid the drying out of the slides at any stage during the washing procedure. Hybridization was performed in the Lucidea Slide Pro station. The slides were removed from the hybridization chambers and placed in a glass jar containing washing solution I (2 x SSC, 0.1 % SDS), pre-warmed to 42°C, then washed 3 x 5 min with this solution by immersing a few times and rocking. Subsequently they were, washed 3 x 5 min with 0.2 x SSC at room temperature followed by wash with water. Each biochip was shortly immersed in a glass jar with isopropanol and dried with a stream of N_2 (see 2.2.5.5).
Slides that were hybridized using the cover-slips were removed from the hybridization chamber and placed in a glass jar containing pre-warmed solution I. Gentle agitation was applied until the cover-slips detached themselves. All additional processing steps were done as described above.

2.2.5.10 Biochips scanning

Two different devices were used for the scanning of the microarray biochips: the ArrayWorx Biochip Reader (Applied Precision) controlled by the ArrayWorx 2.1.2 software suite and the GMS 418 Biochip Reader (Genetic Microsystems) controlled by the Live Image 1.51.0.42 software. Scanning parameters: Scan Area: 39.20 x 20.57 mm, Channel 1: EX 540 / EM 595 nm, 0.100 sec, gain 9.00, Channel 2: EX 635 / EM 685 nm, 0.300 sec, gain 9.00, Resolution: 10.0000 µm.

2.2.6 Peptide SPOT synthesis on cellulose membranes

SPOT synthesis was performed according to Frank, 1992 with the Abimed ASP 222 Automated SPOT Robot. Chemically modified membranes used for peptide synthesis were purchased from AIMS Scientific Products. The APEG ACS01 amino cellulose was used for synthesis of probes covalently bound to this support, subsequently applied for phage affinity selections. The Fmoc-pro-C540 membrane was used for synthesis of soluble N-terminally biotinylated peptides that were cleaved from the solid support and released into the solution

after the synthesis was completed and used for TAIS screens. These peptides were extended with two additional ß-Ala amino acids followed by biotin.

I. Preparation of the Fmoc-pro-C540 support for peptide synthesis:

1. Cleavage of the Fmoc protecting group on the amine of proline by washing the membrane in 20 ml 20 % piperidine in DMF for 5 min on a rocker (under a hood).
2. Washing in 100 ml DMF (1 x 30 s, 4 x 5 min)
3. Washing in 100 ml ethanol (1 x 30 s, 3 x 2 min)
4. Drying in between a folder of 3MM paper with cold air from an airgun.

II. Generation of the SPOTs array

To generate the spots array, the anchor compounds Fmoc-ß-Ala-OH and Boc-Lys[Fmoc]-OH were spotted in the first step on the APEG ACS01 and Fmoc-pro-C540 membranes, respectively, as described in steps below.

1. Preparation of the anchor amino acid derivatives for spotting:

ß-Ala / Ac-Ala (1 : 10)

Fmoc ß-Ala-OH	20 µl	stock solution: 0.2 M amino acid, 0.25 M HOBt in NMP
Ac-Ala-OH	180 µl	stock solution: 0.2 M amino acid, 0.25 M HOBt in NMP
BPB / DMF	4 µl	stock solution: 10 mg BPB in 1 ml DMF
DIC	8 µl	(ca. 0.25 M, 4 µl /100 µl vial)

Vortex, 30 min activation at room temperature

Boc-Lys[Fmoc]

Boc-Lys[Fmoc]-OH	200 µl	stock solution: 0.2 M amino acid, 0.25 M HOBt in NMP
BPB / DMF	4 µl	stock solution: 10 mg BPB in 1 ml DMF

This solution was divided in two vials (100 µl) and activated with 4 µl of DIC, 30 min before the usage.

2. Both membranes were placed on the platform of the synthesizer.
3. Vials containing activated amino acids were placed in a synthesizer rack and a spotting protocol was initiated. The synthesizer dispensed 0.1 µl of activated amino acid solution onto

a single spot on the cellulose membrane. For the Fmoc-pro-C540 membrane the spotting procedure of Boc-Lys (Fmoc) was repeated again after 30 min reaction, in order to obtain better yields.

Both filters, with the arrays of SPOTs, were covered with a glass plate and left for 3 h to complete reaction. Because of the presence of the BPB indicator, all spots were light blue due to their current basic pH. As the reaction proceeded, the spots become yellow indicating formation of non-basic peptide bond.

The acetylation step called “capping” was performed to block all remaining free amino functions between spots. Both membranes were washed separately for 30 s in 100 ml of a capping solution, AC_2O (2 % acetic anhydride in DMF), followed by reaction overnight in 100 ml of a fresh AC_2O on a rocker.

5. Both membranes were washed with 100 ml DMF, (1 x 30 s, 2 x 2 min).
6. The Fmoc amino protecting groups were removed through 5 min reaction in 20 % piperidine / DMF solution.
7. The membranes were washed again with 100 ml DMF (1 x 30 s, 4 x 5 min)
8. The spots were stained light blue by 5 min incubation in 100 ml of 1 % BPB / DMF solution. Repeat if traces of remaining piperidine turn into a dark blue solution.
9. The filters were washed with 100 ml ethanol (1 x 30 s, 4 x 5 min)
10. Dried as described above (I. 4).

III. Assembly of peptides on spots

HOBt-esters of Fmoc amino acid derivatives (2.1.11) must be prepared in NMP for use throughout in spotting reactions. Dissolve 1 mmol of each Fmoc-AA in 5 ml NMP containing 0.25 M HOBt to give 0.2 M Fmoc-AA stock solutions. These are kept in 10 ml plastic tubes that are closed tightly, flash frozen in liquid nitrogen, and stored at -70 C.

1. A set of 0.2M Fmoc amino acid stock solutions was removed from the freezer and thawed to room temperature followed by activation of the desired aliquots with DIC (4 µl / 100 µl of stock aa). The activation reaction was carried out at room temperature for 30 min. Vials with the activated amino acids were placed in the synthesizer rack.
2. The membranes prepared above, were placed onto the platform of the synthesizer for spotting.
3. Solutions of activated amino acids were pre-spotted on a separate 3 MM paper in order achieve regular spot sizes during peptide elongation. 0.2 µl of amino acid solutions were dispensed on a single spot during three consecutive repetitions, each separated by 10 min reaction time. During this step, all spots became yellowish or greenish in color.

The spotted membranes were covered with glass plates and left on the spotter for 2 h or overnight.

All additional steps were performed as described above (see II. 6-12). After completion of the entire process, the membranes were used directly for a next spotting cycle carried out in exactly the same manner or stored in at -20°C until use.
After the final cycle, all peptides were N-terminally acetylated as described in II.6, washed with DMF, ethanol and dried (see above II. 9, 11, 12).

IV. Side-chain deprotection

The deprotection of amino acid side-chains was performed for both membranes according to different procedures.

Immobilized peptides on the APEG amino cellulose membrane:
The dried membrane was immersed in a reaction trough containing 40 ml of deprotection mix (82 % TFA, 10 % DCM, 3 % TIBS, 5 % H_2O, under a hood) and reacted overnight with gentle agitation. The trough was covered very tightly.

Subsequently, the membrane was washed with following solutions (100 ml each):

4 x 2 min DCE
3 x 2 min DMF
2 x 2 min ethanol
3 x 10 min acetic acid (1 M)
4 x 2 min ethanol
The filter was dried with cold air, sealed with a foil and stored at –20°C.

Liquid phase peptides from the Fmoc-pro-C540 membrane:
The dried membrane was immersed in 20 ml of a deprotection mix (50 % CH_2Cl_2, 3 % TIBS, 50 % TFA, 2 % H_2O) for 1 h in a tightly closed trough with gentle agitation (under a hood).
The deprotection mix was exchanged and the reaction was carried out for the next hour.
Subsequently, the membrane was washed with following solutions (100 ml each):

4 x 10 min CH_2Cl_2
3 x 20 min methanol / H_2O + HCl (1+1+0.1 %)
3 x 20 min acetic acid (1 M)

The membrane was dried overnight in a drying chamber under vacuum followed by drying with cold air until the smell of acetic acid was no longer detectable. Peptides were cleaved off from the solid support into the liquid phase according to the following procedure:

1. The identical peptide spots were cut out from the membrane with a scalpel and collected into an Eppendorf tube containing 500 µl of the TAEB (0.1 M) / 20 % ethanol mixture and cleaved off overnight with shaking.
2. The supernatant was transferred in a new Eppendorf tube and the peptide spots were treated with the same mixture again followed by reaction for an additional 2 h with shaking.
3. Supernatants were pooled together and dried in a vacuum centrifuge overnight. Peptides were stored at –20°C.

The purity of peptides was determined with analytical HPLC and their masses were determined by MALDI-MS.

2.2.7 Protein - SPOT peptide binding assay

Binding assays were carried out using spots containing peptides of different length and GST-Mena-EVH1 fusion protein as well as GST protein alone (see section 2.3.5, Fig 2.3.5.25 and Fig 2.3.5.26). Spots were washed sequentially with 70 % ethanol / 1 h, 20 times for 5 min with TBS pH 7.4, at room temperature under agitation on a rocker table and blocked overnight at 4°C with MBS (see Table 2.1.10.1). Next, spots were washed ten times 5 min with TBST pH 7.4 and incubated for 1h at room temperature with 1-2 µg of GST-Mena-EVH1 or 1 µg GST protein followed by TBST wash as above. Two antibodies, primary anti-Mena PAK 73 Ab (dilution 1:1000 and 1:3000) and secondary, POD conjugated anti-rabbit IgG (dilution1:2000) were used in the case of GST-Mena-EVH1 fusion. In the case of GST protein alone, only POD conjugated rabbit anti-GST antibody was used (dilution 1:2000). Spots were incubated for 1 hour at room temperature with corresponding antibodies. Washing procedure with TBST (see above) was repeated after incubation with each antibody. Color reaction was indicating interaction was developed using OPD, substrate solution for peroxidase (see Table 2.1.10.1).

2.2.8 Expression and purification of high-activity *Taq* DNA polymerase

The protocol for DNA *Taq* polymerase purification was adapted from (Pluthero, 1993). The composition of all buffers used here is described in Table 2.1.10.1. An overnight culture containing 100 ml LB medium, 34 µg/ml chloramphenicol and 50 µg/ml kanamycin was initiated with a single colony of BL21 (DE3) pLysS strain containing pET-*Taq* construct and grown at 37°C (200 rpm). The next day, 25 ml of the overnight culture were used to initiate three 800 ml cultures. Bacteria were grown at 37°C with 150 rpm agitation until OD_{600} = 0.4. Production of polymerase was induced by addition of IPTG to a concentration of 1 mM. Bacteria were grown at the same conditions for the additional 5 – 6 h. The cells were harvested by centrifugation at 2800 x g for 15 min at room temperature and the resulting pellet was washed with the buffer A, (100 ml / 1 L culture). The cell recovery was obtained by another 30 min centrifugation step followed by the addition of 50 ml buffer A, re-suspension and the addition of an equal volume of the lysis buffer. Addition of the lysis buffer and sufficient mixing caused cell breaking and the lysate became viscous. Alternatively, the washing step can be omitted and the cell pellet can be re-suspended directly after first centrifugation step in 50 ml of the buffer A. Additionally the cells were efficiently broken by high-pressure homogenization using a French press device, according to manufacturer's guidelines. A glass flask with the lysate was placed in a water bath at 75°C for 1 h. Most of the proteins except thermo-stable DNA *Taq* polymerase got denaturized and fell out of the solution as white flocculent debris. The debris was isolated by centrifugation at 10000 x g, 4°C for 20 min. The yellowish clear supernatant was transferred to a glass flask. The polymerase was recovered at room temperature by gradual addition of well-powdered ammonium sulphate (30 g / 100 ml of supernatant) under rapid stirring. It is very important to add the ammonium sulphate stepwise in small amounts and to dissolve it well before addition of the next portion. The protein precipitate was harvested by centrifugation at 13 000 x g, 4°C for 15 min. The polymerase was visible as a surface precipitate, floating on the surface of the liquid. The supernatant was discarded very gently and the pellet re-suspended well in buffer A, (20 ml / 100 ml of supernatant). The protein was dialyzed for 24 h at 4°C against two changes of the storage buffer (500 ml). The buffer was changed after about 12 h. The dialyzed DNA *Taq* polymerase was stored at –20°C and was ready to use.

3. Results and discussion

3.1 Lambda and T7 phage display systems

3.1.1 Construction of lambda phages presenting model protein domains

The DNA of the phage vector, λDisplay1 was isolated as described in 2.2.3.5. Coding DNA sequences of the three model domains (see Table 1.3.1 in section 1.3), were transferred from the T7Select10-3b constructs (obtained from Krzysztof Bialek, GBF Braunschweig) to the λDisplay1, as shown in Fig 3.1.1. The purified λDisplay1 vector was digested with *Spe*I and *Not*I endonucleases, analyzed by agarose gel electrophoresis for complete cleavage (Fig 3.1.2), and subsequently dephosphorylated followed by ethanol precipitation. The DNA inserts coding for the model domains were PCR amplified. The T7Select10-3b constructs were used as a DNA template for the amplification reaction. Two oligonucleotide primers T7S and T7A were specifically designed for the purpose of PCR (see Table 2.1.6.1.1). The sense primer introduced a *Spe*I restriction site at the 5' end of the insert. The second *Spe*I restriction site, present at the 3' end of the multi cloning site (MCS) of the T7Select10-3b vector, was eliminated by a single T/G point mutation. The *Not*I restriction site was taken from the T7Select10-3b vector (Fig 3.1.1). Before cloning, the PCR amplified sequences were analyzed by agarose gel electrophoresis (Fig 3.1.3). After purification, digestion with both restriction enzymes and ethanol precipitation, appropriate amounts of the inserts and vector DNAs were mixed together in order to perform the ligation reaction as described in (2.2.3.9). The inserted sequences were extended by 21bp and 7bp from the 5' and 3' end respectively in comparison the parent T7Select10-3b constructs. The lambda phages presenting protein domains were assembled using a commercially available packaging kit (see 2.2.3.9). Phage extracts were poured onto the agar plates in order to obtain single well-isolated phage plaques. Twenty phage clones were chosen for PCR analysis to confirm the presence of the DNA sequences coding the model protein domains (Fig 3.1.4). PCR was performed using a universal primer set, designed for the λDisplay1 vector arms (LS and LA) as well as the three primer sets (E1, E2, F1, F2, Y1, Y2), where each set is specific for the respective model domain (see Table 2.1.6.1.1).

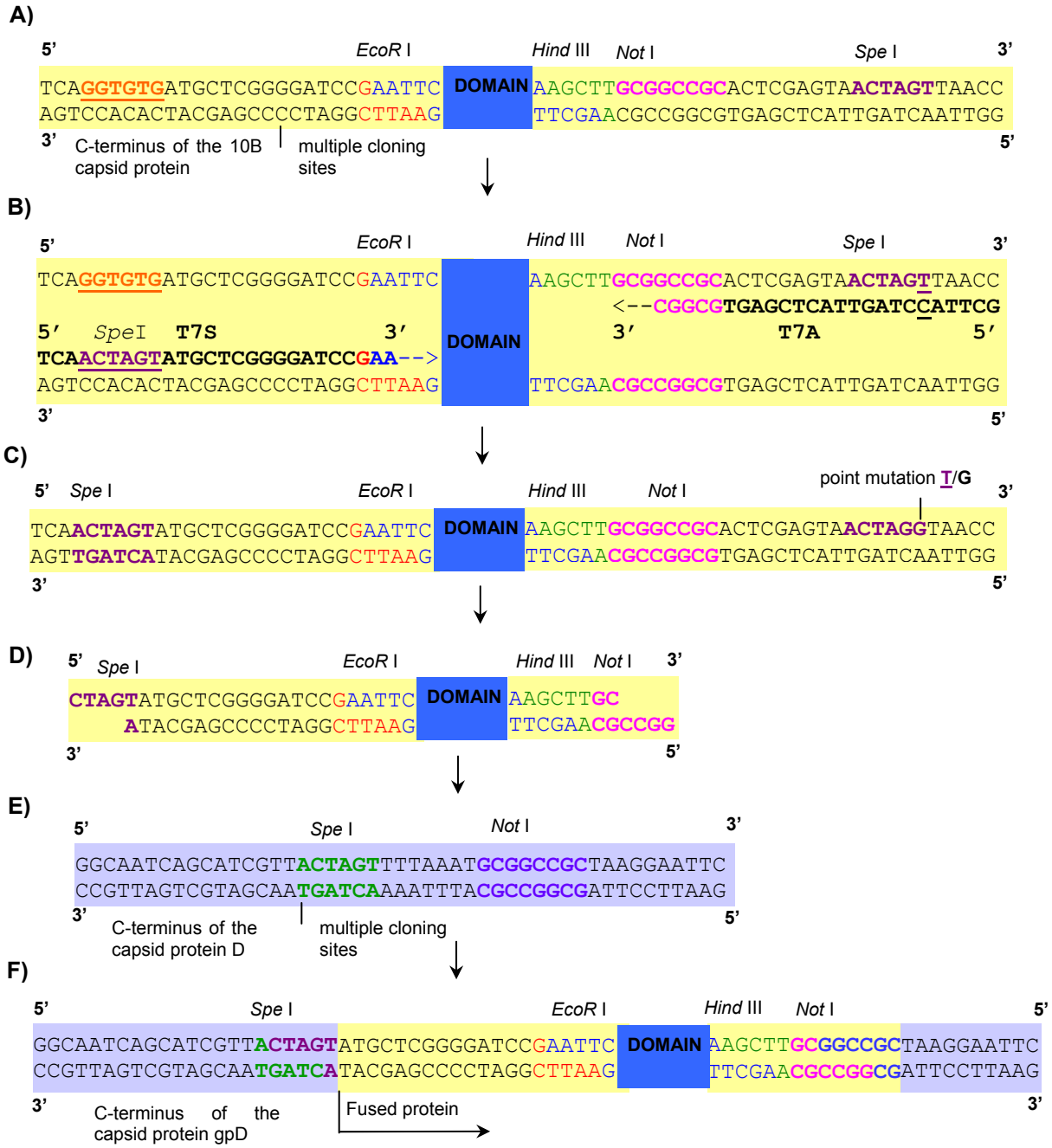


Fig 3.1.1: Cartoon drawing showing the cloning strategy used for the construction of lambda phages presenting model protein domains. **A)** Section of the T7Select10-3B vector and its multiple cloning site with a protein domain (blue) inserted between *EcoR*I and *Hind*III cloning sites. **B)** PCR primers: T7S that introduce a *Spe*I site (purple, underlined) into the appropriate position of the vector (orange, underlined) and T7A that introduce a single point mutation T/G in the second *Spe*I restriction site flanking the 3' end of the insert. **C)** PCR product with the 5' introduced *Spe*I restriction site (purple) and the 3' *Not*I restriction site (pink). **D)** *Spe*I and *Not*I digested PCR product. **E)** Section of the λDisplay1 vector and its MCS. **F)** PCR product cloned between *Spe*I and *Not*I restriction sites at the 3' end of the phage lambda gene D coding the capsid protein gpD.

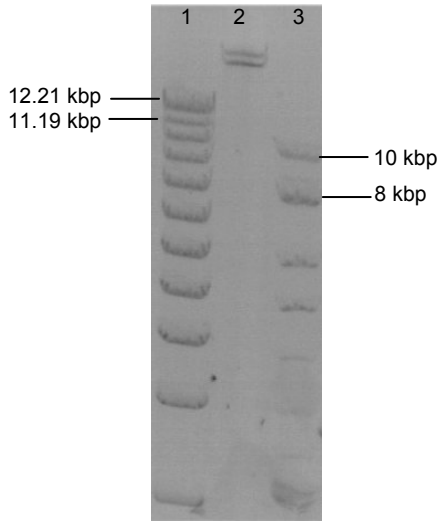


Fig 3.1.2: Ethidium bromide stained 0.7 % AGE analysis of the λDisplay1 vector digested with *Spe*I and *Not*I endonucleases. In lane 1: DNA marker X; 2: *Spe*I and *Not*I digested λDisplay1 vector arms; 3: DNA SMART marker.

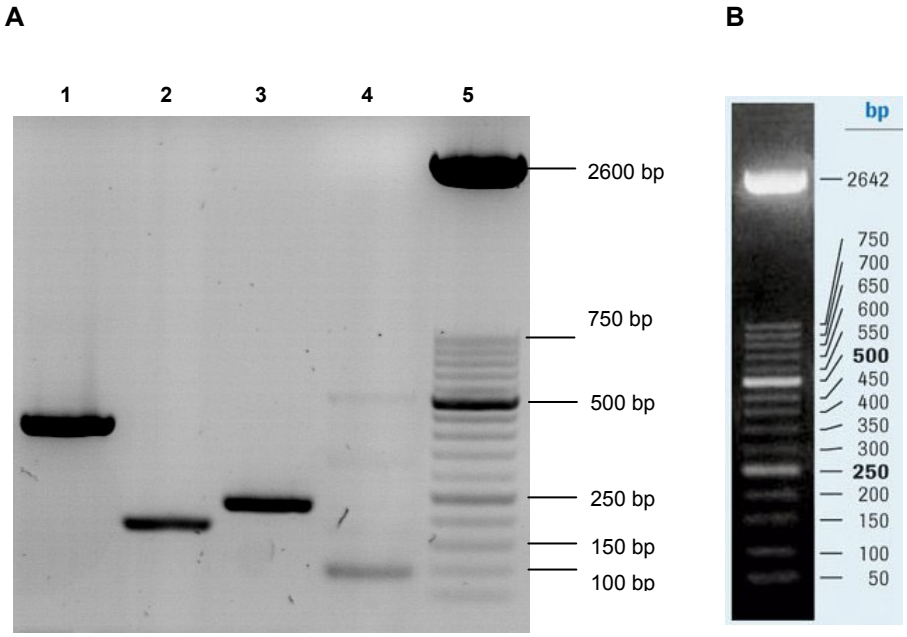


Fig 3.1.3: Ethidium bromide stained 1 % AGE analysis of the PCR products amplified from T7Select10-3b constructs with T7S and T7A primers. **A)** In lane 1: EVH1, 417bp; 2: WW/FE65, 178bp; 3: WW/YAP, 228bp; 4: control PCR sample without DNA template with visible primer oligomerysation products; 5: DNA marker MXIII. **B)** DNA marker MXIII (Boehringer Mannheim).

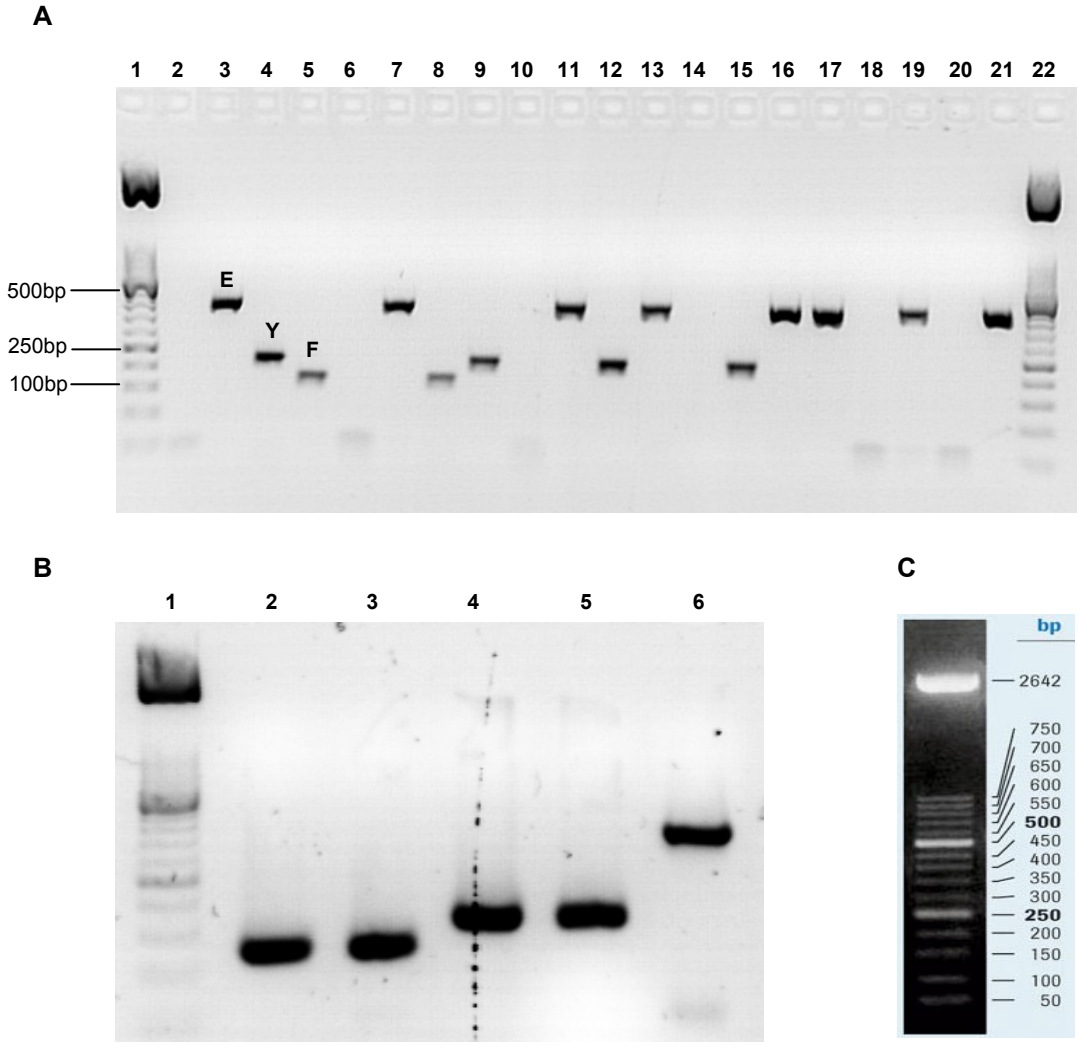


Fig 3.1.4: Ethidium bromide stained 1 % AGE analysis of the PCR products corresponding to the DNA inserts cloned into the λDisplay1 vector. **A)** In line 1: DNA marker MXIII; 2-21: PCR products of 20 randomly chosen phage clones obtained after phage assembly amplified with LS and LA primers (Table 2.1.6.1.1). Expected sizes: (EVH1 (E)-422bp; WW/FE65 (F)-183bp; WW/YAP (Y, 233bp); 22: DNA marker MXIII. **B)** PCR of selected phage clones presenting model domains. Sequences were amplified using domain specific primers. In line 1: DNA marker MXIII; 2-3: WW/FE65, 138bp; 4-5: WW/YAP, 192bp; 5:EVH1, 339bp. **C)** DNA marker MXIII (Boehringer Mannheim).

3.1.2 Comparison of T7 phage with lambda phage

Both phage display systems were then analyzed with respect to their performance in enrichment of domain-presenting phages on the peptide SPOT membranes. The main goal of these studies was a comparison of the efficiencies for the same proteins displayed on both phage types. The phages displaying the three protein domains were used as a model for these studies. Peptide ligands corresponding to these model domains are shown in Table 1.3.1 in section 1.3. Additionally a control peptide that is not a binding ligand for the studied domains (NRPPPAVGPQPAP) was used for monitoring the unspecific phage enrichment. Several different phage mixtures were prepared, that were characterized by high, moderate or low abundance of model phages. Phages displaying model domains were diluted with a mixture of T7 phages displaying IL-4, Fas antigen, ezrin and correspondingly with the wild type lambda phages, lacking an insert. These three T7 phage constructs were engineered previously as described in Bialek *et al*., 2003. Before an appropriate phage mixture was prepared, phage stocks used as a diluent agent, as well as the model phages were poured onto the agar plates and the titers (pfu/ml) were calculated as described in 2.2.4.3. With both types of phages, the same panning conditions were applied except for the elution step, which had to be carried out by direct infection of bacteria in the case of lambda phages due to their instability for elution reagents that were applied for T7 phages (see 2.2.2.4).

Initially, the 1:100 dilutions of phages presenting model domains were prepared (Table 3.1.1). After a single round of panning, on the peace of cellulose membrane containing four domain specific peptide-spots, the PCR and agarose gel electrophoresis were used to identify the number of inserts matching the respective peptide ligands (Fig 3.1.5). The same experiments were carried out twice. The amounts of phages eluted from each peptide SPOT as well as the percentage of correct inserts and the enrichment factors are shown in Table 3.1.2. True enrichment factors were calculated from the ratio of positively verified to the analyzed phages divided by the initial ratio of model to total phages (in this experiment 1:100)

Table 3.1.1: An example showing the preparation of the phage mixture containing the model lambda phages diluted 1:100 in the wild type lambda phages (WT).

Phage displaying domain	Titer pfu/ml	Dilution ratio 1:100	µl
EVH1	$1.9x10^9$	$9.4x10^7$	49.4
WW/FE65	$3.65x10^9$	$9.4x10^7$	25.8
WW/YAP	$6.25x10^9$	$9.4x10^7$	15
WT	$9.4x10^9$	$9.4x10^9$	up to 1000

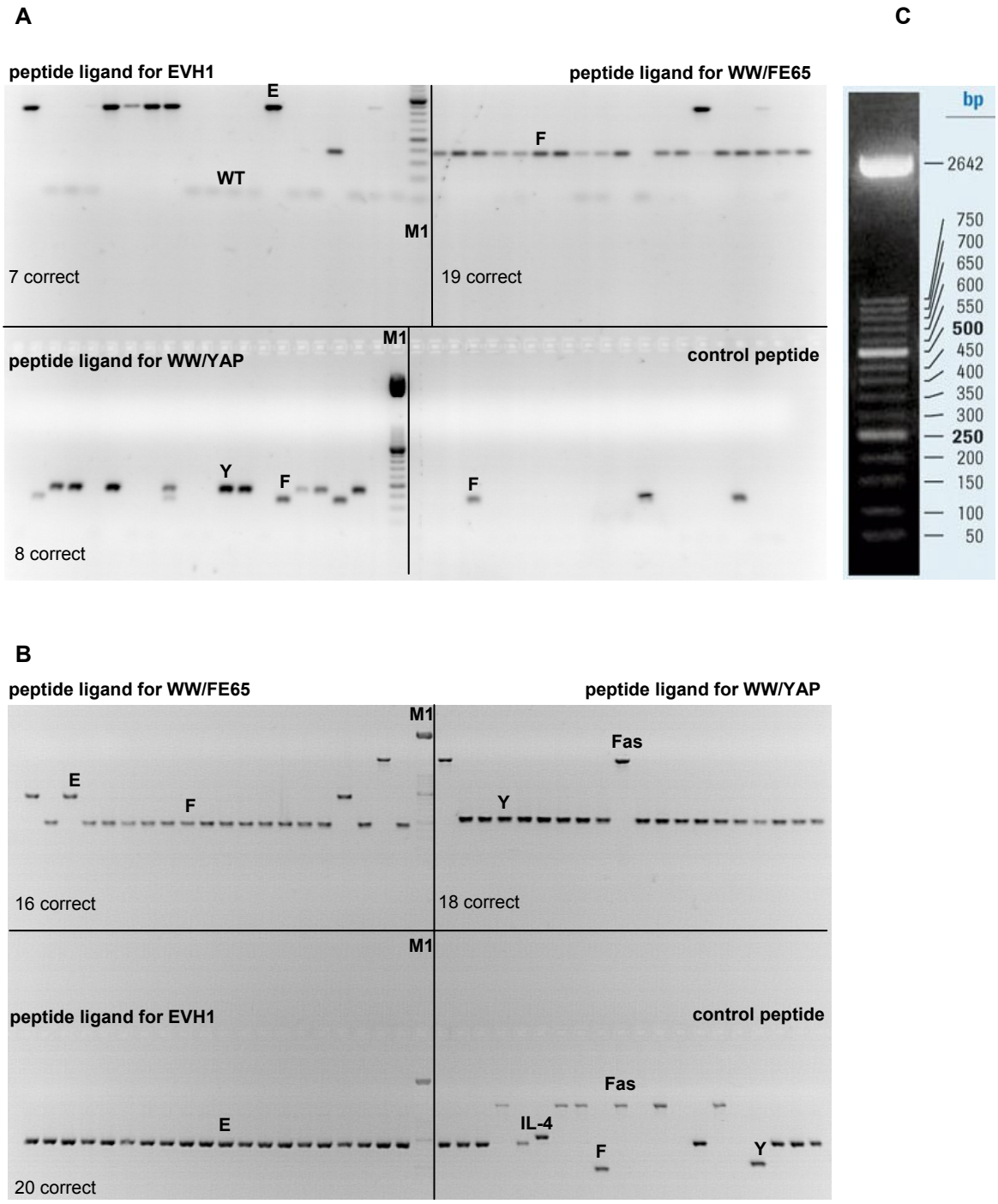


Fig 3.1.5: Ethidium bromide stained 1 % AGE analysis of the PCR products corresponding to the DNA of phages recovered from the peptide-spot ligands. Dilution ratio of model phages: 1:100. Short name code: EVH1 (E), WW/FE65 (F); WW/YAP (Y); Ezrin (Ezr); Interleukin-4 (IL-4),lambda wild type (WT); DNA marker XIII (M1). **A)** Analysis of 20 randomly chosen lambda clones corresponding to each peptide ligand. **B)** Analysis of 20 randomly chosen T7 clones corresponding to each peptide ligand. **C)** DNA marker MXIII (Boehringer Mannheim).

Table 3.1.2: Comparison of the enrichment efficiencies of the lambda and T7 phage display systems, after a single round of panning on membrane bound peptides. Phages presenting model protein domains were diluted 1:100.

Domain	Peptide	Lambda		T7	
		Experiment 1	Experiment 2	Experiment 1	Experiment 2
		Titers of eluted phages			
EVH1	pep-E	$2.8x10^5$	$1.6x10^5$	$3.4x10^4$	$2.1x10^5$
WW/FE65	pep-F	$4.8x10^5$	$4.7x10^5$	$1.4x10^5$	$1.5x10^5$
WW/YAP	pep-Y	$3.1x10^5$	$2.7x10^5$	$1.6x10^5$	$1.5x10^5$
	control	$3x10^4$	$3.6x10^4$	$4.9x10^4$	$1.9x10^4$
		Percentage of correct inserts among 20 analyzed clones [a)]			
EVH1	pep-E	35	50	100	90
WW/FE65	pep-F	95	85	80	80
WW/YAP	pep-Y	40	35	90	95
		Enrichment factors			
EVH1	pep-E	35	50	100	90
WW/FE65	pep-F	95	85	80	80
WW/YAP	pep-Y	40	35	90	95

[a)] 5 % = 1 insert

In the case of the lambda display system, high affinity enrichment to the corresponding peptide SPOTs was observed only for phages presenting WW/FE65 while the two other model phages showed a two or three fold lower representation within the eluted population of phages. On the other hand, the panning experiment with T7 phages indicated very good affinity enrichment for all model phages on their corresponding peptide ligands. However, it was found that the enrichment of the WW/FE65 presenting phages was slightly better in the case of lambda system when compared with the T7 system (see Table 3.1.2).

Since the dilution ratios of model phages applied for the above experiments were relatively low, these phages were already quite enriched within the working phage solution. Therefore, a more realistic dilution ratio ($1:10^4$) was applied for the next experiment that allowed a better investigation of the enrichment efficiencies of both types of phages, especially with respect to the WW/FE65 presenting phages. Panning experiments and clone analysis were performed in the same uniform way as described for the experiments above. The data related to the T7 pannings ($1:10^4$), that were used for comparison of both systems were kindly provided by Krzysztof Bialek (GBF Braunschweig, personal communication), (Bialek *et al.,* 2003;). The agarose gel electrophoresis images are shown in Fig 3.1.6. The enrichment efficiencies of both systems after one round of affinity selection are compared in Table 3.1.3.

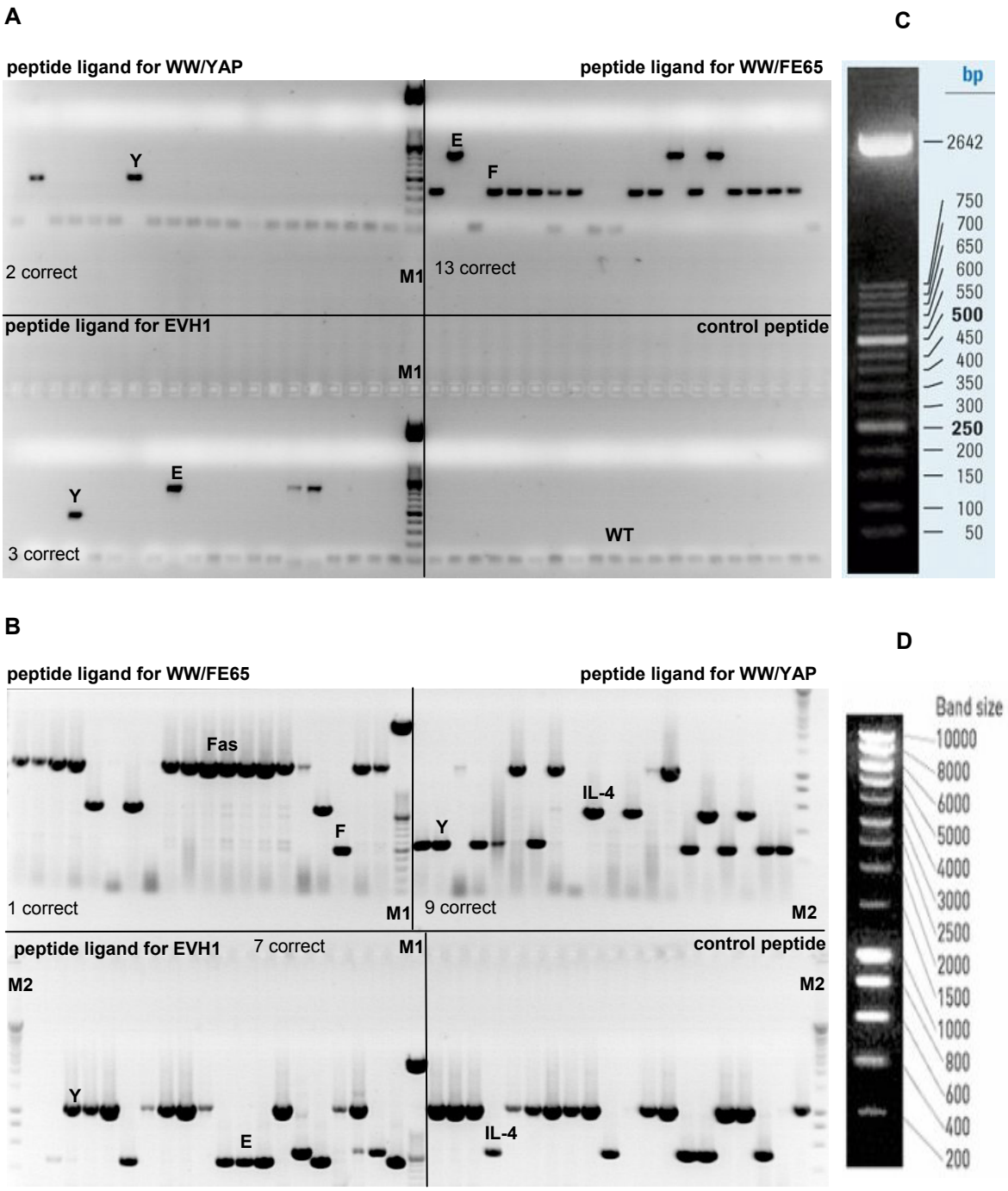


Fig 3.1.6: Ethidium bromide stained 1 % AGE analysis of the PCR products corresponding to the DNA inserts of the phage populations recovered from the peptide SPOT ligands. Dilution ratio of model phages: $1:10^4$. Short name code: EVH1 (E), WW/FE65 (F); WW/YAP (Y); Ezrin (Ezr); Interleukin-4 (IL-4), lambda wild type (WT); DNA marker XIII (M1); DNA marker SMART (M2). **A)** Analysis of 20 randomly chosen lambda clones corresponding to each peptide ligand. **B)** Analysis of 20 randomly chosen T7 clones corresponding to each peptide ligand. The AGE image B), was provided by Krzysztof Bialek). C) DNA marker MXIII (Boehringer Mannheim). D) DNA marker SMART (Eurogentec).

Table 3.1.3: Comparison of the enrichment efficiencies of the lambda and T7 phage display systems, after a single round of panning on membrane bound peptides. Phages presenting model protein domains were diluted 1:10^4

Domain	Peptide	Lambda	T7
		Single experiment	**Average values from two experiments**
		Titers of eluted phages	
EVH1	pep-E	$3.3x10^3$	$6.5x10^3$
WW/FE65	pep-F	$8.4x10^3$	$2.6x10^3$
WW/YAP	pep-Y	$3.2x10^3$	$6.5x10^3$
		Percentage of correct inserts among 20 analyzed clones [a)]	
EVH1	pep-E	15	52.5
WW/FE65	pep-F	65	7.5
WW/YAP	pep-Y	10	44
		Enrichment factors	
EVH1	pep-E	1500	5250
WW/FE65	pep-F	6500	750
WW/YAP	pep-Y	1000	4400

[a)] 5 % = 1 insert

With respect to the lambda display system, results obtained from the experiment that utilized the higher dilution of model phages, confirmed those of the previous investigation. Better enrichment efficiency was observed again for the phages presenting the WW/FE65 model domain. The affinity enrichment of the other two domains was again much less. Contrary to this, in the case of the T7 display system, the WW/FE65 phages once more showed the lowest enrichment efficiency, whereas the two other model phages were represented in relatively high yields with the EVH1 presenting phages as favorites. This difference was not as clear visible after a 1:100 dilution. Obviously these two display systems behave opposite to each other. Additional experiments with the model phages diluted 1:10^6 were performed for the lambda display system. First, only one round of phage panning was carried out. Subsequently the second experiment was done in which the phages enriched after the first round, were propagated and subjected to a second round of selection. The data obtained from both tests were compared with the data acquired from the same type of investigations for the T7 phages, where the model phages were diluted 1:10^5 and 1:10^6 (Bialek *et al.,* 2003). Analysis of 40 (1:10^5) and 20 (1:10^6) phage clones that were randomly chosen from populations eluted from each peptide SPOT (Table 3.1.4) was completed exactly as before (AGE images not shown).

Table 3.1.4: Comparison of the enrichment efficiencies of the lambda and T7 phage display systems, after one and two consecutive rounds of panning on membrane bound peptides. Phages presenting model protein domains were diluted 1:10^5 and 1:10^6 (T7) and 1:10^6 (lambda).

Domain	Peptide	Lambda		T7	
		Single experiment		Average values from two experiments	
		1:10^6		1:10^5	1:10^6
		1 round [a]	2 rounds [a]	1 round [b]	2 rounds [a]
		Percentage of correct inserts among analyzed clones [c]			
EVH1	pep-E	0	25	15	95
WW/FE65	pep-F	10	40	2.5	12.5
WW/YAP	pep-Y	0	0	0	22.5
		Enrichment factors			
EVH1	pep-E	n.d	250 000	15 000	950 000
WW/FE65	pep-F	100 000	400 000	2500	125 000
WW/YAP	pep-Y	n.d	n.d	n.d	225 000

[a] 20 clones analyzed
[b] 40 clones analyzed
[c] 5 % = 1 insert

This last set of experiments addressed the capacity of the method to identify proteins that are only low expressed in the tissue/cells from which the cDNA for the phage library originates. The experiments with the lambda system showed that after a single selection round, only the WW/FE65 domain could be identified, among all verified clones. Its enrichment was four times better than that observed for the corresponding T7 phages isolated also after one panning only (dilution 1:10^5) and comparable to that achieved after two consecutive rounds of T7 selection (dilution 1:10^6). Additionally, when two consecutive lambda phage selection rounds were completed, the abundance of the WW/FE65 module increased four times and was almost four times higher than the abundance of the T7 phages displaying the same domain enriched after two rounds. This confirms once again a better performance of lambda phage system presenting this particular domain in comparison to T7 phages. From the other two domains only the EVH1 could be identified among 20 verified clones achieved after two rounds of panning. Distinct from lambda, the T7 system showed the highest selection performance for the EVH1 domain. It was enriched after a single round of panning and its representation increased evidently after two consecutive pannings, indicating very high enrichment efficiencies of this domain when presented on the T7 phage. The WW domain of YAP was not identified among analyzed clones after one selection round; however, it was sufficiently enriched after two rounds. The comparison of enrichment efficiencies for the model phages, achieved from all experiments performed from both display systems is summarized in Fig 3.1.7.

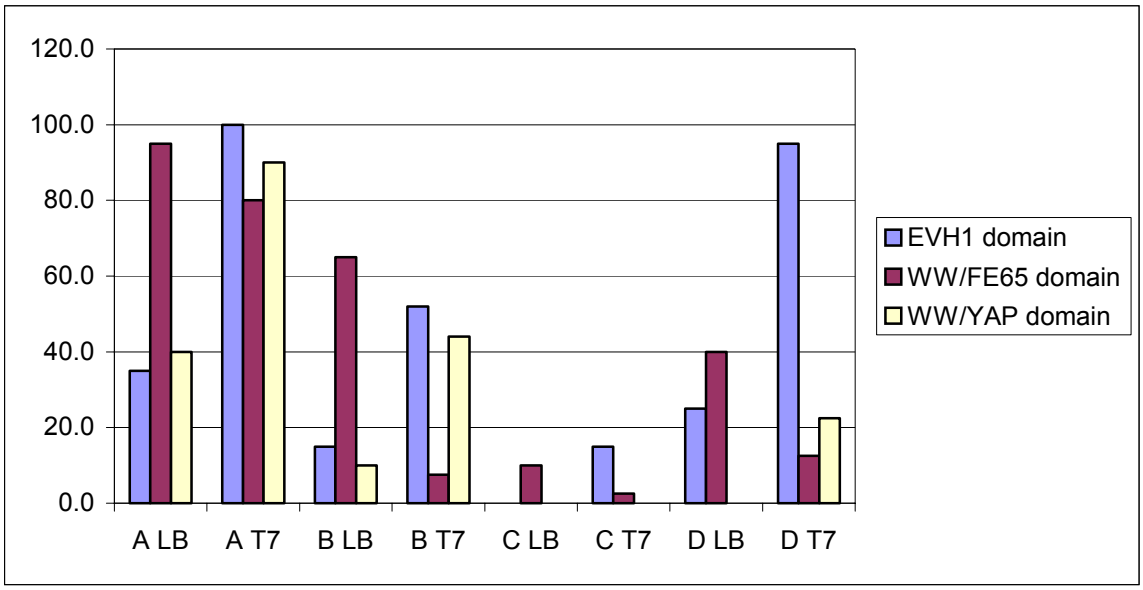


Fig 3.1.7: Affinity enrichment preferences of the lambda (LB) and T7 phages, displaying model protein domains. Comparison of data obtained from four independent experiments utilizing different dilution ratios of model phages: A (1:100); B ($1:10^4$); C ($1:10^5$), one round of panning; D ($1:10^6$), two rounds of panning

The T7Select10-3b display system was designed for middle density display of large proteins with capacities 5-15 copies per phage particle (Rosenberg *et al.*, 1996). Therefore, the T7Select10-3b system is more efficient when strong binders are to be isolated in affinity enrichment experiments. A weak affinity of some proteins must be compensated with higher valence in order to increase the avidity. It was reported that in some cases the display densities of the T7Select10-3b are not enough to achieve the satisfactory phage enrichment (Zucconi *et al.,* 2001). These observations are in agreement with the experimental results described above. In the T7Select10-3b case, the EVH1 domain was always preferentially enriched whereas the WW/YAP domain was less enriched but still better than the WW/FE65 domain. The observed enrichment efficiencies correlate well with the binding constants of the corresponding protein-peptide complexes, 5 µM and 54 µM for the EVH1 and WW/YAP respectively. A binding constant for the WW/FE65 complex is not reported. Its behaviour however, with reference to the other two model domains indicates a weaker affinity. The λDisplay1 system was designed for high copy display of proteins (Castagnoli *et al.,* 2001). Its high display density (up to 405 copies) was advantageous for affinity isolation of weaker binders that couldn‘t be efficient isolated with the T7 system, as reported by Zucconi *et al.*, 2001.

Contrary to the T7 system, the data obtained from all experiments clearly revealed the highest enrichment efficiency for the WW/FE65 when displayed by lambda phages. According to the observations reported for low affinity proteins (Zucconi *et al.,* 2001), it was deduced that the WW/FE65 domain belongs most probably to the weaker binding domains. Its high enrichment efficiency was a result of the avidity increase caused by a high-density display. Surprisingly, the enrichment of the two other modules was much less as compared to the T7 system. Why were the EVH1 and WW/YAP model phages, despite their higher affinities for the peptide ligands, less represented within the recovered lambda phage populations? It could be suggested that the high display valence strongly increased their overall binding avidity to the high-density peptide SPOTs. As a result of the strong binding, these two model phages may not be sufficiently infective during the applied elution procedure based on direct infection of bacteria. However, the chance that the bacteria reach the peptide bound phage and become infected should be similar for all three phages displaying model domains. An explanation for this experimental result remains obscure. It can be only tentatively suggested that some structural changes important for infection are required during the docking process of bacteria to phage and that the increased binding affinity of the phage to its peptide SPOT does not allow for these changes. An experimental proof for this speculation could be reached by utilizing lower densities of peptides on the membrane spots in order to decrease the binding avidity of these phages. An alternative explanation for the lower enrichment of the EVH1 and WW/YAP domains could be an inefficient folding of these domains when fused to the C-terminus of the lambda capsid protein gpD, causing reduced affinities to their peptide ligands. This may be corroborated by e.g. plasmon resonance measurements (BiaCore) with immobilized peptides and the domain-presenting phages compared to the free protein domains. Although the lambda display system has proven to be preferable when weak protein-peptide interactions are to be identified, the overall performance of the T7 display system proves to be more reliable and general. Additionally, the higher stability against different elution reagents, shorter life cycle, more convenient handling, as well as the availability of many commercial cDNA libraries are other advantages that support the choice of the T7 display system for the goal of this thesis.

3.2 Accessible surface scanner software tool

The objective of the work in this section was an evaluation of a computer software tool that was developed for prediction of surface-exposed protein regions based on a published algorithm (see section 1.2.1). This tool was subsequently utilized for peptide generation. Several model protein sequences were analyzed that contain continuous epitopes, which were proven to be ligands for diverse protein modules. The resulting output files were examined for the presence of these epitopes. Additionally, selected sequences that belong to proteins of known structure were scanned with the software tool and the output peptides were checked for their appearance in the surface-exposed regions, deduced from the protein structure.

3.2.1 Predictions of surface probability for model ligands

The preliminary tests were performed with four proteins that are well-characterized, natural binding partners for the model protein domains studied in this thesis: the WBP1 and WBP2, ligands of the WW/YAP, as well as Mena and ActA, interacting with the WW/FE65 and EVH1 respectively. The initial parameters applied for the following peptide selections were: peak height threshold 2.0, minimum width 1.0 and smoothing 2.0. The peptide length was defined as 25 aa with the overlap of 10 aa concerned for those surface regions, which were longer than 25 aa. The original sequences and the output peptide files, as well as surface probability plots for these proteins are shown in Fig 3.2-1 – 3.2-4. Comparison of the input and output sequences showed that the motifs responsible for interaction with their corresponding modules were present among selected peptide fragments in the all four cases. Moreover, the overall, possible number of peptides that would be required for a systematic sequence scanning was reduced by about 70 % after a computational analysis with the ASS software (see Table 3.2.1).

Table 3.2.1: Comparison of results obtained after computer analysis with the ASS software and the systematic peptide scan of the ActA, Mena, WBP1 and WBP2 protein sequences.

Protein domain	Binding protein	Size (aa)	Binding core motifs	Total No. of motifs	No. of motifs in the output file	No. of peptides (systematic scan) 25 aa	No. of peptides (ASS scan) 25 aa
EVH1	ActA	639	FPPPP/ PPIP	4	3	42	12
WW FE65	Mena	802	PPLP	6	4	54	15
WW YAP	WBP1	304	PPPPY	1	1	20	6
WW YAP	WBP2	261	PPPPY	3	3	17	5

Additionally, for the ActA, WBP1 and WBP2, the residues that were scored as surface-probable were found directly within the cores of binding motifs, whereas such residues were located outside the motif in the case of Mena (see A1-D1, turquoise).

A1

spP33379 **ActA INPUT SEQUENCE**

```
MGLNRFMRAMMVVFITANCITINPDIIFAATDSED
SSLNTDEWEEEKTEEQPSEVNTGPRYETAREVSSR
DIKELEKSNKVRNTNKADLIAMLKEKAEKGPNINN
NNSEQTENAAINEEASGADRPAIQVERRHPGLPSD
SAAEIKKRRKAIASSDSELESLTYPDKPTKVNKKK
VAKESVADASESDLDSSMQSADESSPQPLKANQQP
FFPKVFKKIKDAGKWVRDKIDENPEVKKAIVDKSA
GLIDQLLTKKKSEEVNASDFPPPPTDEELRLALPE
TPMLLGFNAPATSEPSSFEFPPPPTDEELRLALPE
TPMLLGFNAPATSEPSSFEFPPPPTEDELEIIRET
ASSLDSSFTRGDLASLRNAINRHSQNFSDFPPIPT
EEELNGRGGRPTSEEFSSLNSGDFTDDENSETTEE
EIDRLADLRDRGTGKHSRNAGFLPLNPFASPVPSL
SPKVSKISAPALISDITKKTPFKNPSQPLNVFNKK
TTTKTVTKKPTPVKTAPKLAELPATKPQETVLREN
KTPFIEKQAETNKQSINMPSLPVIQKEATESDKEE
MKPQTEEKMVEESESANNANGKNRSAGIEEGKLIA
KSAEDEKAKEEPGNHTTLILAMLAIGVFSLGAFIK
IIQLRKNN
```

;spP33379 **ActA OUTPUT PEPTIDES**

```
.pep
TDEWEEEKTEEQPSEVNTGPRYETA (40-64)
LKEKAEKGPNINNNNSEQTENAAIN (93-117)
SAAEIKKRRKAIASSDSELESLTYP (141-165)
DSELESLTYPDKPTKVNKKKVAKES (156-180)
QLLTKKKSEEVNASDFPPPPTDEEL (250-274)
PATSEPSSFEFPPPPTDEELRLALP (290-314)
ATSEPSSFEFPPPPTEDELEIIRET (326-350)
FTDDENSETTEEEIDRLADLRDRGT (409-433)
KNPSQPLNVFNKKTTTKTVTKKPTP (479-503)
ATKPQETVLRENKTPFIEKQAETNK (515-539)
QKEATESDKEEMKPQTEEKMVEESE (551-575)
RSAGIEEGKLIAKSAEDEKAKEEPG (585-609)
```

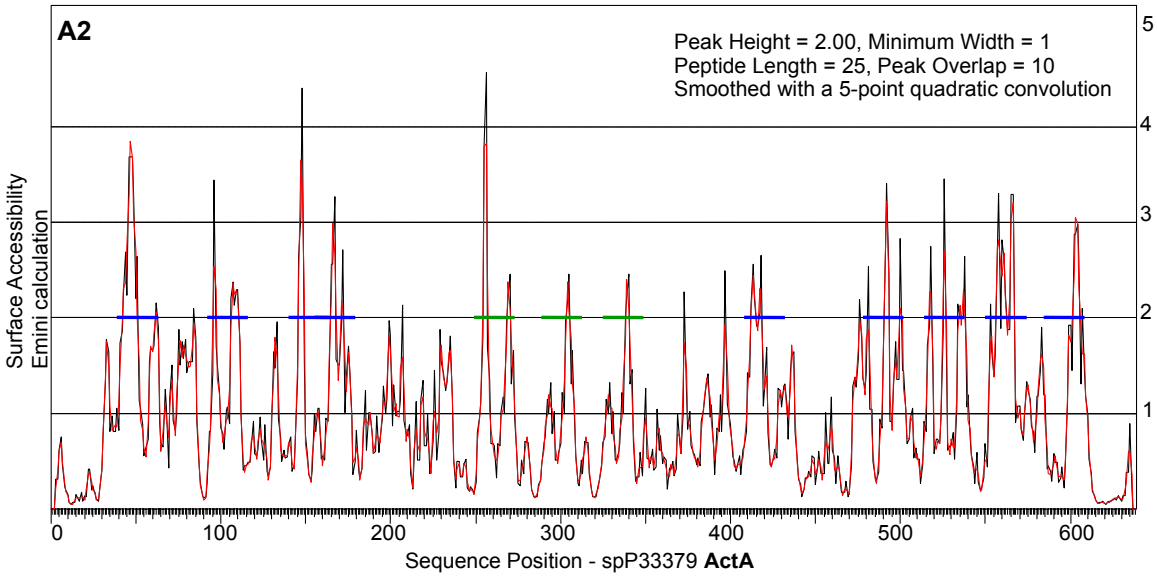


Fig 3.2.1: A computer analysis of the ActA protein sequence with the ASS software tool. **A1)** Comparison of the input protein sequence with the output file showing selected peptides. Green: bolded sequences indicate binding motifs for the EVH1 domain. Turquoise: residues determined to be surface-exposed, **A2)** Surface probability plot. Horizontal lines illustrate selected protein fragments that were used for generation of peptides. Green lines indicate the binding motifs. Selection parameters: peak height threshold 2.0, minimum width 1.0 and smoothing 2.0, peptide length 25 aa.

B1

spP97764 **WBP1 INPUT SEQUENCE**

MARASSRNSSEEAWGSLQAPQQQQSPAASSLEGAI
WRRAGTQTRALDTILYHPQQSHLLRELCPGVNTQP
YLCETGHCCGETGCCTYYYELWWFWLLWTVLILFS
CCCAFRHRRAKLRLQQQQRQREINLLAYHGACHGA
GPVPTGSLLDLRLLSAFKPPAYEDVVHHPGT**PPPP**
YTVGPGYPWTTSSECTRCSSESSCSAHLEGTNVEG
VSSQQSALPHQEGEPRAGLSPVHIPPSCRYRRLTG
DSGIELCPCPDSSEGEPLKEARASASQPDLEDHSP
CALPPDSVSQVPPMGLASSCGDIP

;spP97764 **WBP1 OUTPUT PEPTIDES**

.pep

RASSRNSSEEAWGSLQAPQQQQSPA (3-27)
IWRRAGTQTRALDTILYHPQQSHLL (35-59)
FRHRRAKLRLQQQQRQREINLLAYH (110-134)
F**K**PPAY**E**DVVHH**P**GT**PPPP**Y**T**VGPG (157-181)
HQEGEPRAGLSPVHIPPSCRYRRLT (220-244)
PDSSEGEPLKEARASASQPDLEDHS (255-279)

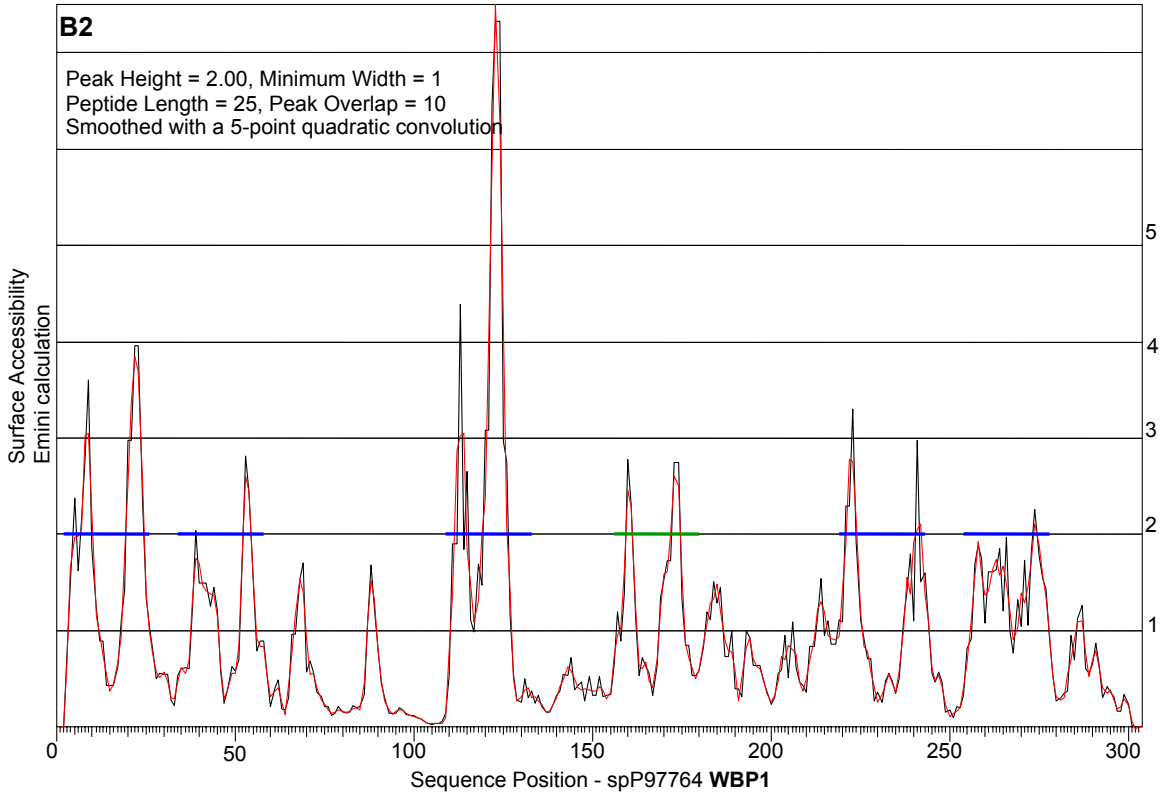


Fig 3.2.2: A computer analysis of the WBP1 protein sequence with the ASS software tool. **B1)** Comparison of the input protein sequence with the output file showing selected peptides. Green: bolded sequence indicates binding motifs for the WW/YAP domain. Turquoise: residues determined to be surface-exposed. **B2)** Surface probability plot. Horizontal lines illustrate selected protein fragments that were used for generation of peptides. Green lines indicate binding motifs. Selection parameters: peak height threshold 2.0, minimum width 1.0 and smoothing 2.0, peptide length 25 aa.

C1

```
spP97765 WBP2 INPUT SEQUENCE
MALNKNHSEGGGVIVNNTESILMSYDHVELTFNDMKNVP
EAFKGTKKGTVYLTPYRVIFLSKGKDAMQSFMMPFYLMK
DCEIKQPVFGANFIKGIVKAEAGGGWEGSASYKLTFTAG
GAIEFGQRMLQVASQASRGEVPNGAYGYPYMPSGAYVFP
PPVANGMYPCPPGYPYPPPPPEFYPGPPMMDGAMGYVQP
PPPPYPGPMEPPVSGPSAPATPAAEAKAAEAAASAYYNP
GNPHNVYMPTSQPPPPPYYPPEDKKTQ
```

```
;spP97765 WBP2 OUTPUT PEPTIDES
.pep
LNKNHSEGGGVIVNNTESILMSYDH (3-27)
TFNDMKNVPEAFKGTKKGTVYLTPY (31-55)
GMYPCPPGYPYPPPPPEFYPGPPMM (162-186)
YVQPPPPPYPGPMEPPVSGPSAPAT (192-216)
PGNPHNVYMPTSQPPPPPYYPPEDK (234-258)
```

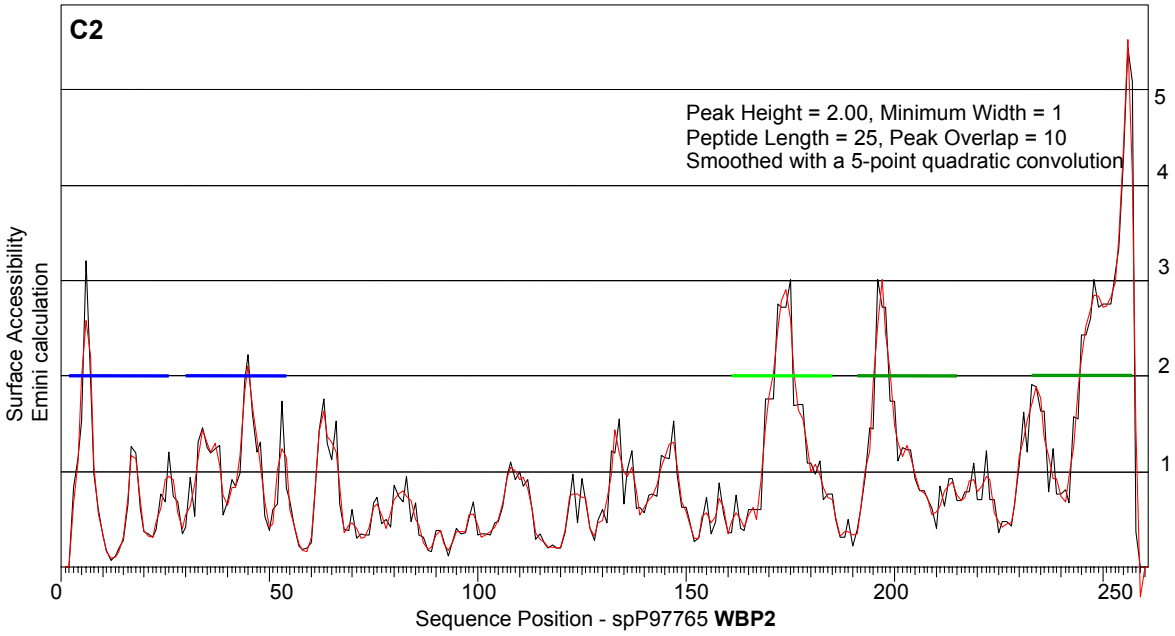


Fig 3.2.3: A computer analysis of the WBP2 protein sequence with the ASS software. **C1)** Comparison of the input protein sequence with the output file showing selected peptides. Green: bolded sequences indicate motifs that bind to the WW/YAP domain. Turquoise: residues determined to be surface-exposed, dark green: peptide in the forward orientation (PPPY), light green: peptide in the reverse orientation. **C2)** Surface probability plot. Horizontal lines illustrate selected protein fragments that were used for generation of peptides. Green lines indicate binding motifs in forward (dark) and reverse (light) orientations. Selection parameters: peak height threshold 2.0, minimum width 1.0 and smoothing 2.0, peptide length 25 aa.

D1

spQ03173 **Mena INPUT SEQUENCE**

```
MSEQSICQARAAVMVYDDANKKWVPAGGSTGFSRV
HIYHHTGNNTFRVVGRKIQDHQVVINCAIPKGLKY
NQATQTFHQWRDARQVYGLNFGSKEDANVFASAMM
HALEVLNSQEAAQSKVTATQDSTNLRCIFCGPTLP
RQNSQLPAQVQNGPSQEELEIQRRQLQEQQRQKEL
ERERMERERLERERLERERLERERLEQEQLERQRQ
EREHVERLERERLERLERERQERERERLEQLEREQ
VEWERERRMSNAAPSSDSSLSSAPLPEYSSCQPPS
APPPSYAKVISAPVSDATPDYAVVTALPPTSTPPT
PPLRHAATRFATSLGSAFHPVLPHYATVPRPLNKN
SRPSSPVNTPSSQPPAAKSCAWPTSNFSPLPPSPP
IMISSPPGKATGPRPVLPVCVSSPVPQMPPSPTAP
NGSLDSVTYPVSPPPTSGPAAPPPPPPPPPPPPPP
PLPPPPLPPLASLSHCGSQASPPPGTPLASTPSSK
PSVLPSPSAGAPASAETPLNPELGDSSASEPGLQA
ASQPAESPTPQGLVLGPPAPPPPPPLPSGPAYASA
LPPPPGPPPPPPLPSTGPPPPPPPPPLPNQAPPP
PPPPPAPPLPASGIFSGSTSEDNRPLTGLAAAIAG
AKLRKVSRVEDGSFPGGGNTGSVSLASSKADAGRG
NGPLPLGGSGLMEEMSALLARRRRIAEKGSTIETE
QKEDRNEDAEPITAKAPSTSTPEPTRKPWERTNTM
NGSKSPVISRPKSTPSSQPSANGVQTEGLDYDRLK
QDILDEMRKELAKLKEELIDAIRQELSKSNTA
```

```
;spQ03173 Mena OUTPUT PEPTIDES
.pep
DDANKKWVPAGGSTGFSRVHIYHHT (17-41)
CGPTLPRQNSQLPAQVQNGPSQEEL (135-159)
VQNGPSQEELEIQRRQLQEQQRQKE (150-174)
QLQEQQRQKELERERMERERLERER (165-189)
MERERLERERLERERLERERLEQEQ (180-204)
LERERLEQEQLERQRQEREHVERLE (195-219)
QEREHVERLERERLERERQERE (210-234)
RLERERQERERERLEQLEREQVEWE (225-249)
QLEREQVEWERERRMSNAAPSSDSS (240-264)
TVPRPLNKNSRPSSPVNTPSSQPPA (342-366)
PPTSGPAAPPPPPPPPPPPPPLP (434-458)
PGPPPPPPLPSTGPPPPPPPPPLP (565-589)
PPPPPPPPLPNQAPPPPPPPPAPPL (580-604)
PPPPPPAPPLPASGIFSGSTSEDNR (595-619)
ARRRRIAEKGSTIETEQKEDRNEDA (685-709)
AKAPSTSTPEPTRKPWERTNTMNGS (714-738)
WERTNTMNGSKSPVISRPKSTPSSQ (729-753)
EGLDYDRLKQDILDEMRKELAKLKE (762-786)
```

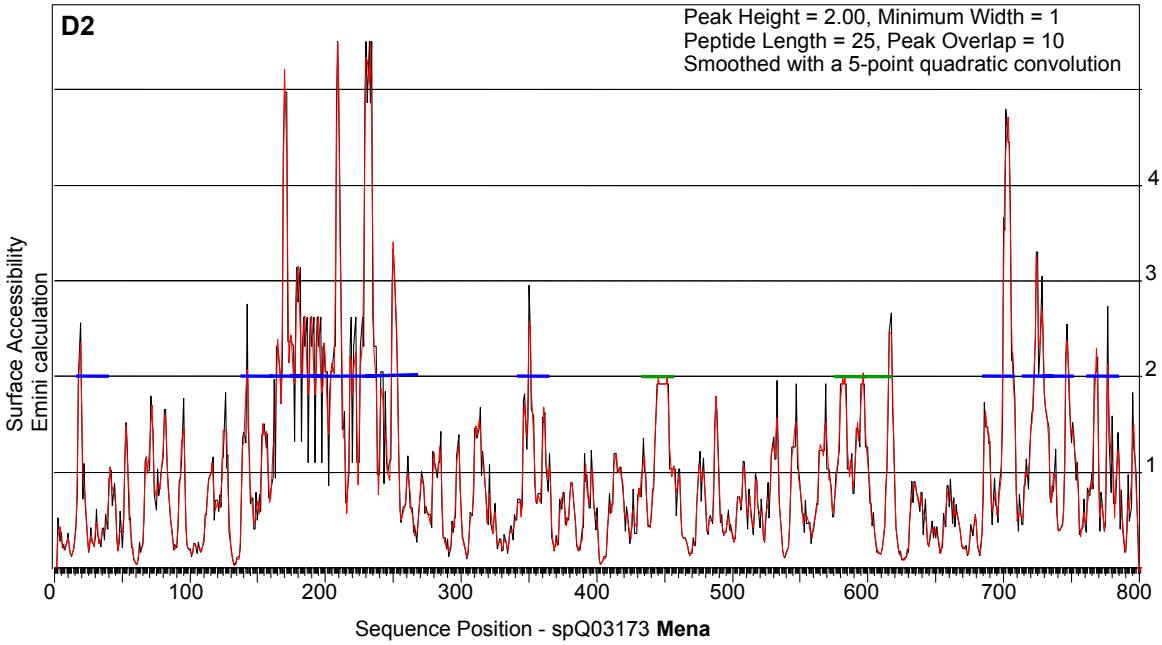


Fig 3.2.4: A computer analysis of the Mena protein sequence with the ASS software. **D1)** Comparison of the input protein sequence with the output file showing selected peptides. Green: sequences indicate motifs that bind to the WW/FE65 domain. Turquoise: residues determined to be surface-exposed, bold: core of the binding motifs, normal: consecutive polyproline stretches preceding the core motifs, underlined: not selected motifs **D2)** Surface probability plot. Horizontal lines illustrate selected protein fragments that were used for generation of peptides. Green lines indicate binding motifs. Selection parameters: peak height threshold 2.0, minimum width 1.0 and smoothing 2.0, peptide length 25 aa.

To investigate the influence of parameter variations on the surface probability prediction process, additional studies were carried out using the above protein sequences. Only one out of three parameters was altered before each scanning procedure was initiated. Their values ranged from 1.5 to 3.5 for peak height threshold, from 1.0 to 6.0 for minimum width and from 0 to 5.0 for smoothing, (see Table 3.2.2).

Table 3.2.2: Influence of the variable selection parameters on the number of peptides and binding motifs in the output files. Bold: variable parameter, blue: parameters and results from the previous analysis, red: values obtained without smoothing function.

Results								**Selection parameters**		
No. of peptides in the output file				**No. of binding motifs in the output file**				**Peak height**	**Min. width**	**Smoothing**
ActA	WBP1	WBP2	Mena	ActA	WBP 1	WBP 2	Mena			
21	7	6	20	4	1	3	4	**1.5**	1.0	2.0
12	6	5	15	3	1	3	3	**2.0**	1.0	2.0
9	5	4	8	1	1	3	0	**2.5**	1.0	2.0
6	2	2	7	1	0	2	0	**3.0**	1.0	2.0
3	2	1	5	1	0	1	0	**3.5**	1.0	2.0
12	6	5	15	3	1	3	3	2.0	**2.0**	2.0
12	6	5	14	3	1	3	3	2.0	**4.0**	2.0
11	5	4	13	3	1	3	3	2.0	**6.0**	2.0
17	7	5	13	4	1	3	1	2.0	1.0	**0**
13	5	4	13	3	1	3	3	2.0	1.0	**3.0**
12	5	5	10	3	1	3	2	2.0	1.0	**4.0**
10	4	3	11	1	1	3	2	2.0	1.0	**5.0**

Variation of the peak height threshold parameter revealed having a major effect on the amounts of generated peptides and preservation of the proper binding motifs in the output files. Its increase caused a dwindling of the peptide quantities for all studied proteins, but frequently at the expense of the correct ligands. For the ActA and Mena, rise of the peak height threshold from 1.5 to 2.0, indicated a loss of only one binding motif, however, this became critical when further heightened from 2.0 to 2.5, causing a reduction of the majority or all binding motifs respectively. In the case of the WBP1 and WBP2, their core motifs responsible for binding to WW domains were located within higher scored regions, as deduced from the surface probability plots (Fig 3.2.2 and Fig 3.2.3). Therefore, elimination of binding motifs in these cases was observed only when the peak height threshold value was increased up to 3.0 or higher (Table 3.2.2). Apart from the peak height threshold, the other two parameters didn't seriously affect surface predictions for these two proteins. The

omission (0) or application (2.0) of the smoothing function however, revealed to have different influence on the ActA and Mena (Table 3.2.2, blue and red). This omission was advantageous for the ActA, resulting in selection of the fourth binding sequence (FPPIP), which was also found when the peak height threshold was reduced to 1.5. The total number of generated peptides however, was higher in this case than that recorded for the analysis with the smoothing 2.0. A detailed investigation of surface probabilities (Fig 3.2.5, A1, A2), revealed that the non-smoothed value for residue H373 (2.269) was higher than the defined peak height threshold (2.0). Therefore a fragment containing this residue was scored as a surface-probable. Nevertheless, this value turned out to be smaller (1.739) than the defined threshold, when calculated with the smoothing 2.0, which resulted in elimination of the entire fragment from the output file. An opposite effect was observed for Mena protein, where smoothing (2.0) resulted in an increase of the number of binding core motifs from one to four. Analysis of surface-probability values for residues from an example fragment (F1) of Mena revealed a slight increase of the value for residue P452 (from 1.927 to 2.005), when smoothing 2.0 was applied (Fig 3.2.5, B1, B2). This qualified the P452 as a surface exposed and promoted peptide generation process.

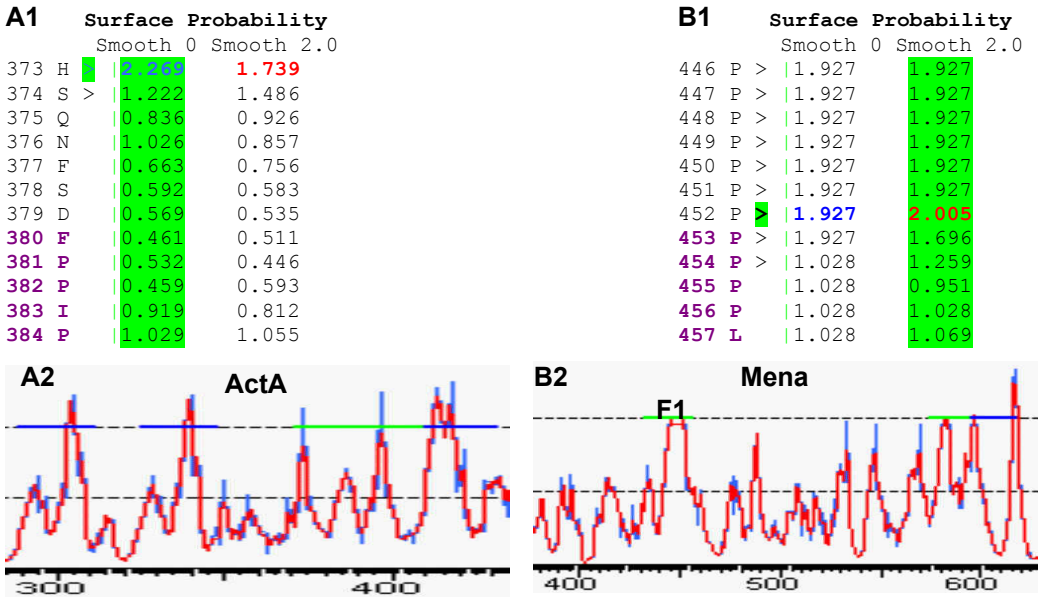


Fig 3.2.5: The influence of smoothing function on fragment generation process. Selection parameters: peak height threshold 2.0, minimum width 1.0, peptide length 25 aa. **A1)** Section of ActA protein sequence. **B1)** Section of Mena protein sequence. Surface probability values of the residues H373 and P542: blue bold-no smoothing, red bold-smoothing 2.0. Violet bold: binding motifs, green: selected region. **A2** and **B2)** Surface probability plots for the ActA and Mena. Blue plot: (no smoothing), red plot (smoothing 2.0). Horizontal lines: blue; regions selected with smoothing parameters 2.0 and 0.0, green; regions selected with smoothing (ActA) or without smoothing (Mena).

The F1 region contains a strain of eight proline residues having the same surface probability value (1.927), which is only slightly smaller than defined peak height threshold (2.0). Despite the entire F1 region was not scored as surface probable when scanned without smoothing and above parameters, it could be considered as moderately surface exposed and would be definitely selected if lower peak height threshold was employed.

3.2.2 Predictions of surface probability for different domain-ligands

The above investigations demonstrated that variations of the peak height threshold and activation or deactivation of the smoothing function had somewhat different effects on diverse protein sequences. The need for uniform selection parameters that would be beneficial for generation of relatively low quantities of peptides while preserving the occurrence of protein binding motifs was a goal of the auxiliary tests. In order to evaluate the accuracy of the prediction with the ASS software, additional 21 proteins, which interact with the SH3, GYF, EH and PDZ domains were examined. In a similar fashion, the resulting files were checked for occurrence of the appropriate binding motifs. Initially, 25mer peptides were generated and the peak height threshold (2.0) and smoothing (2.0) were applied, as they provided satisfactory results in the previous calculations. Subsequently, the same calculations were carried out with reduced peak height threshold (1.5), in order to verify whether this change would reinstate those motifs that were lost after the first scanning run. The results of these calculations are shown in Table 3.2.3. All tests revealed that the predominance of motifs relevant for binding was preserved after surface probability analysis carried out with the more restrictive peak height threshold (2.0). Nevertheless, for some isolated proteins (AFAP, Abl, Sos-1, Shb), this computing approach appeared to be less beneficial, suffering up to 50 % loss of known epitopes. In addition, multiple NPF binding motifs present in the Hrb (4) and Epsin-1 (3), were not found in the resulting files. However, all these motifs for the Hrb and one for the Epsin-1 were reinstate after reduction of the peak height threshold to 1.5. This change restored also a minimal core sequence for the fourth motif in the SOS-1 protein, which was only partially selected when the peak height threshold was 2.0. These three proteins were the only examples showing an improvement with reference to the number of the desired motif sequences in the output files, as a response to the peak height threshold reduction. Analogously, as observed for the previous calculations, reduction of this threshold resulted in an increase of total amounts of output peptides, for all studied proteins. General difficulties were observed with recovery of two PDZ binding motifs located at the C-terminus of the ARVCF and ErB4 proteins using the above as well as several other parameter variations. However, a similar motif from the Frizzled protein that is located internally was successfully selected even after scanning with the most restrictive

thresholds. Difficulties with the positive selections of sequences located at the N- and C-termini are related to the algorithm performance. To calculate a surface probability value for a certain amino acid (n) the values of surrounding amino acids (n-2 and n+3) are necessary. Unfortunately, such calculations are not possible for the first 2 and last 3 residues. Surface probability values for these terminal residues are always equal 0, (Emini *et al.,* 1985). Therefore motifs binding to PDZ domains, which include the last three C-terminal residues couldn't be positively scored by the ASS tool (see 3.2.5).

Table 3.2.3: Comparison of the ASS selected binding motifs with motifs present within input sequences. The protein domains, their interaction partners and binding motifs, as well as numbers of motifs and peptides present in the output files are shown. Selection parameters: peptide length 25 aa, peak height thresholds 2.0 and 1.5, minimum width 1.0, smoothing 2.0. The quantities of peptides obtained for peak height thresholds are shown. Green: binding motifs selected by the ASS tool with the peak height threshold 2.0, black: full motifs or their fragments that were not selected via the software tool, blue: residues from the binding motifs, which were selected with the peak height threshold 1.5, bold: key residues necessary for interaction. Motifs with missing key residues were not accepted as valuable binding partners.

Protein/ Domain	Binding protein	Binding motifs (aa)	No. of motifs nature/ASS		No. of output peptides		Reference
			P.H [a)] 2.0	P.H [a)] 1.5	P.H [a)] 2.0	P.H [a)] 1.5	
Src/SH3 class I	Hs AFAP	PPQMP**L**P**E**I**P**QQ PPDNGP**P**PL**P**TS	2/1	2/2	20	26	Flynn *et al.,* 1993
	Hs CDC42 GAP	TAPKPM**P**PR**P**PL	1/1	1/1	7	7	Barfod *et al.,* 1993
Src/SH3 class II	Hs hnRNP K	PLP**P**PP**P**PR**G**G	1/1	1/1	8	12	Weng *et al.,* 1994
	Hs Dynamin	GGA**P**PV**P**S**R**PG GPP**P**QV**P**S**R**PN RAP**P**GV**P**S**R**SG	3/3	3/3	20	24	Gout *et al.,* 1993
	Hs PI3K p85	QPA**P**AL**P**P**K**PP	1/1	1/1	19	25	Kapeller *et al.,* 1994
	Hs Shb	GGP**P**PG**P**G**R**RG TKS**P**PQ**P**P**R**PD	2/1	2/1	14	19	Karlsson *et al.,* 1995
Crk/N SH3	Hs Abl	QA**P**EL**P**T**K**TR VS**P**LL**P**R**K**ER EK**P**A**L**P**K**RA	3/2	3/2	30	39	Ren *et al.,* 1994
	Hs C3G	PP**P**A**L**P**K**KR TP**P**A**L**P**K**KR KP**P**P**L**P**K**KN PP**P**A**L**P**K**QR	4/4	4/4	26	35	Knudsen *et al.,* 1994

Protein/ Domain	Binding protein	Binding motifs (aa)	No. of motifs nature/ASS		No. of output peptides		Reference
			P.H [a] 2.0	P.H [a] 1.5	P.H [a] 2.0	P.H [a] 1.5	
Grb2/N SH3 class II	Hs SOS-1	PVPPPVPPRRR DSPPAIPPRQP ESPPLLPPREP IAGPPVPPRQS	4/2	4/3	34	44	Rozakis-Adcock *et al.,* 1993
	Rn Synapsin-1	NLPEPAPPRPS PPGPAGPIRQA	2/2	2/2	24	31	McPherson *et al.,* 1994
Endophilin/ SH3	Synaptojanin-1	PKRPPPPR	1/1	1/1	39	58	Zucconi *et al.,* 2001
Eps15/EH	NUMB	NPF	1/1	1/1	13	21	Salcini *et al.,* 1997
	Hrb	4 x NPF	4/0	4/3	9	21	Yamabhai *et al.,* 1998; Lucke *et al.,* 2005
	Epsin-1	3 x NPF	3/0	3/1	8	19	Chen *et al.,* 1998
	Synaptojanin-1	3 x NPF	3/3	3/3	39	58	Haffner *et al.,* 1997
CD2BP2/ GYF	CD2 SmB	SHRPPPPGHRV GMPPPGMR RPPPPGMR	1/1 2/2	1/1	6 4	12 8	Kofler *et al.,* 2004
Lin1p/GYF	Prp-8	PPPPGF PPPPGY PPPPGL	3/3	3/3	55	71	Bialkowska *et al.,* 2002
Erbin/PDZ	ARVCF	DSWV- C term	1/0	1/0	27	40	Laura *et al.,* 2002
PSD95/ PDZ	ErbB-4	TVV- C term	1/0	1/0	30	55	Garcia *et al.,* 1999
Dvl1/PDZ	Frizzled	GKTLQSWRRFYH internal PDZ motif	1/1	1/1	12	16	Wong *et al.,* 2003

[a] Peak height threshold

In the second stage of these studies, the same 21 protein sequences were subjected for analysis using various parameters. Initially, the same calculations as above with the peak height threshold 2.0 were carried out with and without smoothing in order to investigate its influence on the process of peptide selection (Table 3.2.4). Similarly as observed previously for the ActA and Mena, activation or deactivation of smoothing showed opposite effects on the Dynamin-1 and Hrb proteins. The Dynamin-1 contains three poly-proline motifs for the Src/SH3 domain. All of them were found within the resulting fragments when smoothing 2.0 was applied, whereas only two were hit upon its omission (pink). Opposite to this, none of the four NPF motifs were obtained in the case of the Hrb protein employing smoothing;

however two were recovered after scanning without this function (red). The total number of generated sequences was higher for the experiments performed without smoothing parameter, which is in agreement with previous observations. Apart from the Dynamin-1 and Hrb, no other protein showed a response to smoothing variations in this approach.

In order to test whether the same motifs will be preserved after reduction of peptide length, all 21 proteins were analyzed using the following parameters: peptide length 15 aa combined with the peak height thresholds 2.0 and 1.5, with and without smoothing. Loss of several appropriate motifs was observed in many cases (see Table 3.2.4). Obviously, surface probability values calculated for residues from these motifs were lower than the respective peak height thresholds. Since their incorporation into generated peptide depends on its defined length, those motifs, which are more distant from the residues scored as surface-probable were eliminated upon peptide shortening. As expected, this elimination was more frequent in the case of calculations with the peak height threshold 2.0. Nevertheless, several of those lost motifs could be reinstated when lower peak height threshold (1.5) was employed. Only in one case (SOS-1), the shortening of peptide length resulted in scoring an additional binding motifs as surface probable (Table 3.2.4 blue). It was observed for all proteins that the overall number of peptides in the output files systematically increased (up to 50 %) after reduction of each parameter. The parameters chosen for future predictions are required to provide a minimal number of peptides that have to be chemically synthesized, with possible high preservation of the appropriate binding motifs in the resulting file. In this case larger numbers of proteins could be spotted in the form of peptides on a single membrane allowing for global analysis of protein-peptide interactions. Since the application of the peak height threshold 2.0 in combination with the peptide length of 25 aa provided satisfactory results with reference to the preservation of binding motifs and total amounts of selected peptides, these two values appeared to be the most appropriate for future calculations. Both, omission and application of smoothing will be still considered in the future studies.

Table 3.2.4: Quantitative comparison of the ASS selected binding motifs and amounts of peptide fragments generated with the ASS using different parameters. Peak height threshold 2.0 and 1.5, minimum width 1.0, smoothing 2.0 and 0.0, peptide length 25 and 15 aa.

Domain binding protein	Nature	No. of binding motifs						No. of output peptides					
		P.H [a] 2.0 (25 aa)		P.H [a] 2.0 (15 aa)		P.H [a] 1.5 (15 aa)		P.H [a] 2.0 (25 aa)		P.H [a] 2.0 (15 aa)		P.H [a] 1.5 (15 aa)	
		Smooth		Smooth		Smooth		Smooth		Smooth		Smooth	
		0	**2.0**	**0**	**2.0**	**0**	**2.0**	**0**	**2.0**	**0**	**2.0**	**0**	**2.0**
AFAP	2	1	1	0	0	1	1	23	20	32	29	49	46
CDC42 GAP	1	1	1	1	1	1	1	7	7	7	8	10	8

Domain binding protein	Nature	No. of binding motifs						No. of output peptides					
		P.H [a] 2.0 (25 aa)		P.H [a] 2.0 (15 aa)		P.H [a] 1.5 (15 aa)		P.H [a] 2.0 (25 aa)		P.H [a] 2.0 (15 aa)		P.H [a] 1.5 (15 aa)	
		Smooth		Smooth		Smooth		Smooth		Smooth		Smooth	
		0	**2.0**	**0**	**2.0**	**0**	**2.0**	**0**	**2.0**	**0**	**2.0**	**0**	**2.0**
hnRNP K	1	1	1	1	1	1	1	11	8	18	17	28	28
Dynamin	**3**	**2**	**3**	**1**	**1**	**1**	**1**	25	20	30	26	43	38
PI3K p85	1	1	1	1	1	1	1	24	19	39	32	45	41
Shb	2	1	1	1	1	1	1	18	14	21	18	30	25
Abl	3	2	2	1	1	2	2	27	30	46	39	66	44
C3G	4	4	4	4	3	4	4	29	26	41	37	51	47
SOS-1	**4**	**2**	**2**	**4**	**3**	**4**	**3**	37	34	67	55	85	82
Synapsin-1	2	2	2	1	1	1	1	23	24	32	27	42	38
Synaptojanin	1	1	1	1	1	1	1	48	39	63	60	92	81
NUMB	1	1	1	1	0	1	1	17	12	33	22	40	37
Hrb (NPF)	**4**	**2**	**0**	0	0	**2**	**2**	14	9	16	10	30	29
Epsin-1 (NPF)	3	0	0	0	0	1	0	13	8	15	13	27	26
Synaptojanin	3	3	3	2	1	2	2	48	39	63	60	92	81
CD2	1	1	1	1	0	1	1	8	6	14	11	25	22
SmB	2	2	2	2	2	2	2	4	4	5	5	10	7
Prp-8	3	3	3	2	2	3	3	57	55	101	85	123	126
ARVCF C-term	1	0	0	0	0	0	0	33	27	46	35	60	51
ErbB-4 C-term	1	0	0	0	0	0	0	40	39	66	52	89	80
Frizzled	1	1	1	1	1	1	1	12	12	16	17	26	23

[a] Peak height threshold

3.2.3 ASS predictions versus protein structure

Future attempts to evaluate the ASS performance were concerned with a comparison of results achieved for a single sub-unit of the pentameric C-Reactive Protein (CRP), (Shrive *et al.,* 1996, PDB 1GNH), with its crystal structure. Presence of several surface exposed loops, which are a potential source of linear epitopes as well as interesting for analysis β-structures and α-helices within the CRP sub-unit, were the main reasons for choosing this protein as a model. Calculations were performed with and without smoothing of surface probability values, using the following parameters; peptide length 25 aa, peak height threshold 2.0, minimum peak width 1.0, overlap 10 aa. Peptide fragments acquired from the computer analysis were verified for their appearance in the regions that are exposed on the protein

surface. The surface probability plots on Fig 3.2.6, show protein regions predicted as a surface exposed (horisontal lines) as well as peaks representing residues having the highest surface probabilities. The red colored plot, corresponds to predictions with smoothing 2.0. The blue plot shows a surface probabilitiy profile generated without application of this function (0). Four 25 aa peptides were generated from the predicted regions in both cases (1-4, blue lines). An additional region represented by a single, C-terminally located peptide 5 (green line), was scored as a surface probable, only when smoothing was avoided. A detailed analysis of this fragment revealed that the surface probability value of the highest scored residue (L190) was higher than the defined peak height threshold (2.120 > 2.0), resulting in its positive selection. In contrast, activation of smoothing function reduced this value drastically to 1.515, eliminating this fragment from the output peptide file (Fig 3.2.7). Additionally, all residues in this region were much less scored than residues in other selected regions, as deduced from the surface probability plots (Fig 3.2.6). Therefore, the entire C-terminal region of the CRP monomer should be considered here, less surface-probable.

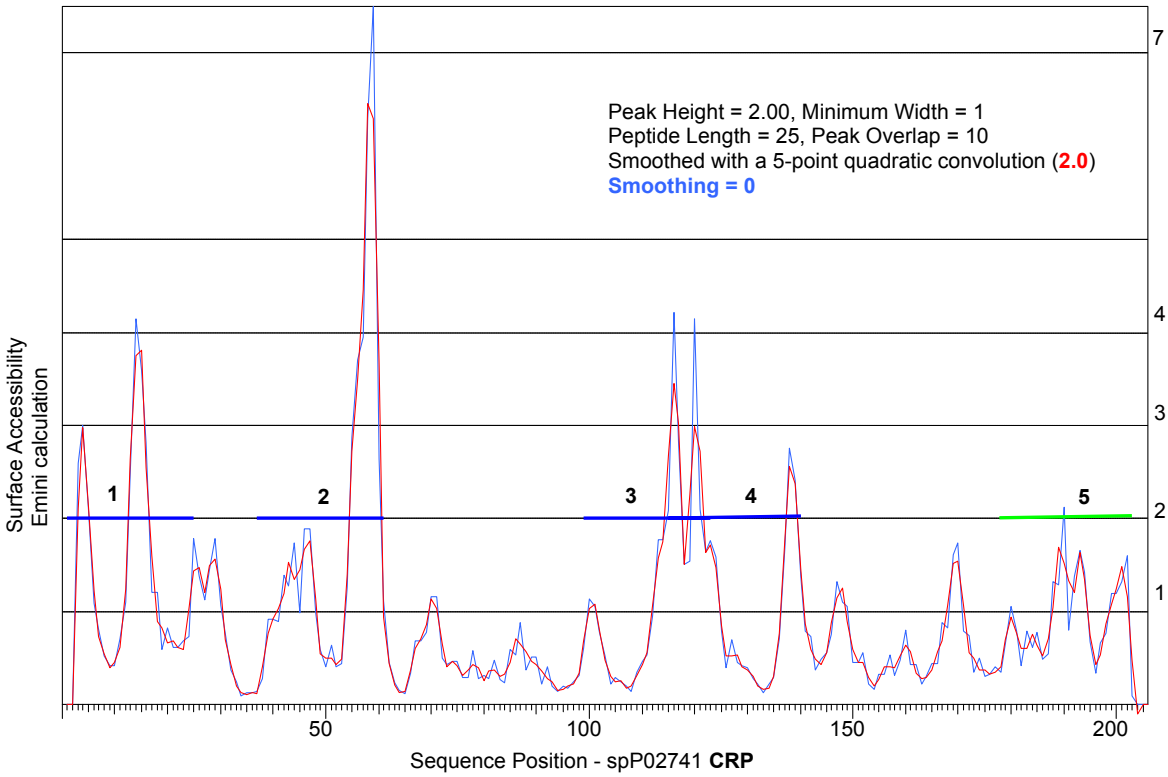


Fig 3.2.6: Surface probabilities prediction for the CRP. Selection parameters: peak height threshold 2.0, minimum width 1.0, smoothing 2.0 and 0, overlap 10 aa, peptide length 25 aa. Superimposed surface probability plots. Light blue: smoothing = 0, red: smoothing 2.0. Horizontal lines indicate the ASS generated 25 aa peptides: blue with smoothing 2.0, green with smoothing 0.

Surface Probabilities

peak 2.0, pep 25 aa

Smooth:		**0**		**2.0**
178 G		|0.347		0.412
179 P		|0.650		0.695
180 F		|1.057		0.942
181 S		|0.793		0.769
182 P		|0.423		0.603
183 N		|0.785		0.603
184 V		|0.616		0.758
185 L		|0.780		0.644
186 N		|0.490		0.526
187 W		|0.545		0.708
188 R	>	|1.321		1.046
189 A	>	|1.287		1.689
190 L	>	**|2.120**	25aa	**1.515**
191 K		|0.803		1.337
192 Y		|1.377		1.207
193 E		|1.653		1.639
194 V		|1.431		1.347
195 Q		|0.678		0.738
196 G		|0.339		0.435
197 E		|0.659		0.537
198 V		|0.761		0.872
199 F		|1.189		1.076
200 T		|1.189		1.236
201 K		|1.321		1.490
202 P		|1.604		1.162

Fig 3.2.7: ASS calculations for the C-terminal sequence of the CRP scanned with and without smoothing. Abbreviations: “>” if determined to be on the surface, “I” if included in a generated peptide fragment, arrows: directions of the fragments elongation. Bold: surface probability values for L190.

As next, the accuracy of the above calculations was checked by comparison of results with the crystal structure of CRP sub-unit. The first four generated peptides were found within the surface exposed loops in the protein structure (see Fig 3.2.8, yellow, LP: 1- 4). All residues having the highest surface probability values (red) were located within these loops, documenting the accuracy of performed predictions. According to the protein structure, the C-terminal region (R1) is evidently surface exposed. However, the ASS predictions didn’t reveal such high surface probability for this region, as described for the peptide-5 above. Moreover, two additional regions, the α-helix (H) and β-sheet (βa), which seems to be displayed on the protein surface, were not scored as surface-probable by the software. It can be speculated that predictions of such secondary structures is more difficult for the ASS, than predictions of extended loop regions. This was supported by the analysis of two additional protein sequences (Nef-HIV1 and Crk-SH2) and comparison of acquired predictions with their crystal structures (data not shown). Nevertheless, several exposed loops were scored as surface probable in both cases. Since the linear binding motifs for various protein domains are often located in such loops, generation of peptides from these

regions is highly desirable. Therefore the general performance of the ASS software tool is satisfactory.

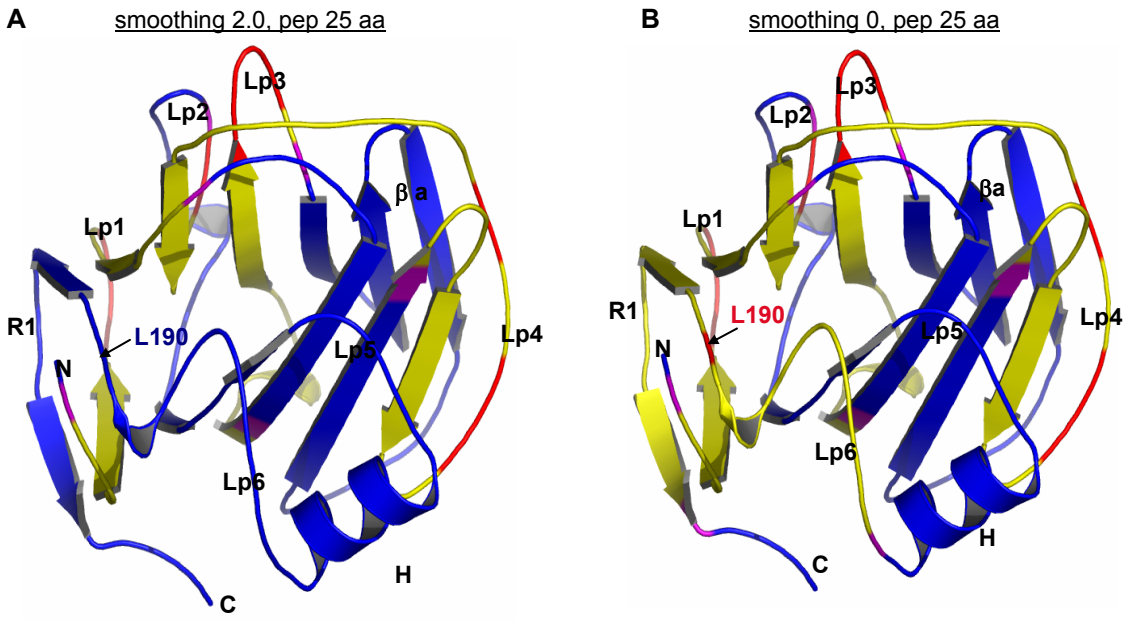


Fig 3.2.8: Comparison of the ASS software predictions with the structure of CRP monomer (Shrive *et al*., 1996). Selection parameters: peak height threshold 2.0, minimum width 1.0, smoothing 2.0 (**A**), no smoothing (**B**), overlap 10 aa. Abbreviations: yellow: superimposed peptide fragments, red: residues showing the highest surface probabilities, pink: termini of generated peptides, blue: regions that were not selected via the ASS, Lp: loop, H: helix, R: region, β: β-sheet.

3.2.4 Comparison of the surface probability and solvent accessibility algorithms

The final investigations were concerned with comparison of performances of the surface probability predicting algorithm (ASS) with the algorithm predicting solvent accessibility (software PH-Dacc), which was previously used by another group for studies on the p53 and Znf 217 proteins (Pavlik *et al.,* 2003). Using the PH-Dacc software, Pavlik and colleagues selected four and two epitopic peptides for both proteins, respectively. These were experimentally verified for binding to antibodies that were enriched before from a phage display library with the protein as bait. The main goal of studies performed here was a test whether the same epitopes could be successfully predicted by the ASS software. Both protein sequences were scanned using the peak height threshold 2.0, minimum peak width 1.0, peptide lengths 25 aa, with and without smoothing of surface values. As an example, the

ASS surface probability plots, achieved here for the p53 protein sequence were compared with a plot showing solvent accessibility predictions for the same protein (Fig 3.2.9). From several algorithms tested by Pavlik, the solvent accessibility was shown as the best performance. Therefore the second chain flexibility algorithm used in his studies, which is also visible on Fig 3.2.9-A, will not be compared here with the surface probability algorithm. Despite visible similarities, several significant differences between both plots were observed including regions (R1-R3, pink) containing the four binding epitopes.

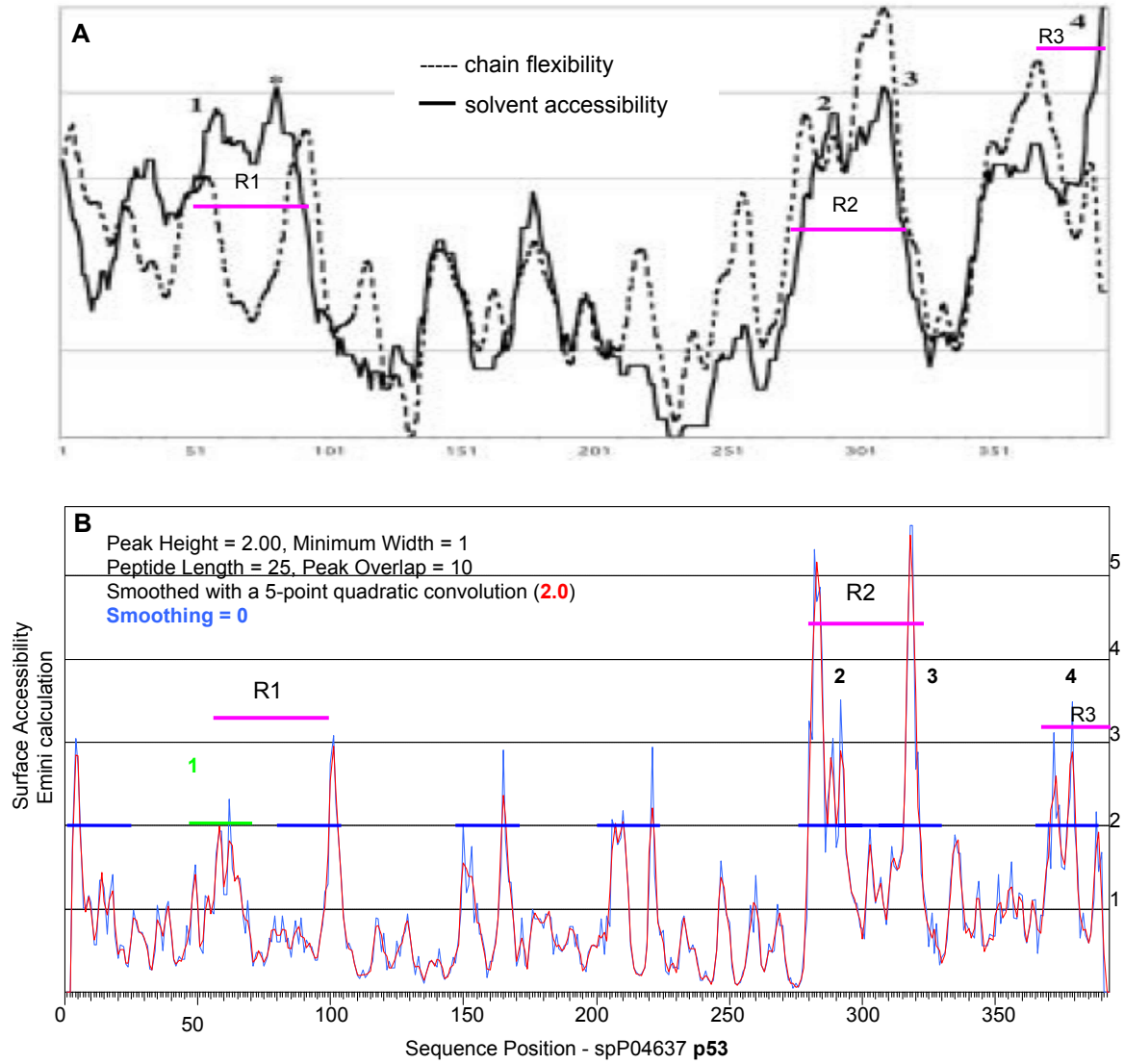


Fig 3.2.9: Comparison of plots achieved for the p53 protein after calculations using the solvent accessibility and the surface probability algorithms. **A**) Chain flexibility and solvent accessibility plots (Pavlik *et al.,* 2003, modified). **B)** Surface probability plots. Scanning parameters: peak height threshold 2.0, minimum width 1.0, smoothing 2.0 (red) and 0 (blue), overlap 10 aa, peptide length 25 aa. Horizontal lines indicate the ASS generated peptides (blue, smoothing = 2.0, green: smoothing = 0). Regions R1-R3, show differences in probability calculations between different algorithms (pink).

Then, the peptides acquired using the ASS software, were checked for the presence of the epitopes described by Pavlik. Scanning of p53 protein sequence using smoothing of surface probability values, resulted in generation of three peptides containing the epitopic sequences see (Table 3.2.5). Omission of smoothing resulted in the recovery of peptide containing the fourth epitope. Nevertheless, only one fragment (3) contained a full length binding motif. The remaining three sequences were up to three residues shorter than the corresponding peptides selected with the PH-Dacc software. Additional calculations were performed using the same parameters with peptides length of 15 aa. This reduction resulted in even stronger shortening of the desired epitopic sequences for all peptides with exception of the peptide No.3 (see Table 3.2.5), indicating that the majority of peptides selected as solvent accessible by Pavlik, were not fully scored as surface probable with the ASS. These sequences were however incorporated into the 25 aa long fragments due to peptide extension process.

Table 3.2.5: Comparison of peptides achieved from the p53 sequence after application of algorithms calculating surface probability (ASS) and solvent accessibility (Pavlik *et al.*, 2003). Several different selection parameters applied for the ASS calculations are shown. Red: epitopic residues, which were predicted with the ASS, blue: epitopic residues, which were not predicted with the ASS.

p53 protein	
Surface probability algorithm (ASS)	**Solvent accessibility algorithm**
pep 25 aa, peak 2.0	
smoothing 2.0	
1 ------------------------	**1** GPDEAPRMPEAAPPV 59-73
2 KGEPHHELPPGSTKRALPNNTSSSP (292-316)	**2** KKGEPHHELPPGSTK 291-305
3 ALPNNTSSSPQPKKKPLDGEYFTLQ (307-331)	**3** PNNTSSSPQPKKKPL 309-323
4 SSHLKSKKGQSTSRHKKLMFKTEGP (366-390)	**4** RHKKLMFKTEGPDSD 379-393
without smoothing	
1 SPDDIEQWFTEDPGPDEAPRMPEAA (46-70)	**1** GPDEAPRMPEAAPPV 59-73
2 GEPHHELPPGSTKRALPNNTSSSPQ (293-317)	**2** KKGEPHHELPPGSTK 291-305
3 LPNNTSSSPQPKKKPLDGEYFTLQI (308-332)	**3** PNNTSSSPQPKKKPL 309-323
4 SSHLKSKKGQSTSRHKKLMFKTEGP (366-390)	**4** RHKKLMFKTEGPDSD 379-393
pep 15 aa, peak 2.0	
smoothing 2.0	
1 ---------------	**1** GPDEAPRMPEAAPPV 59-73
2 RTEEENLRKKGEPHH (283-297)	**2** KKGEPHHELPPGSTK 291-305
3 PNNTSSSPQPKKKPL (309-323)	**3** PNNTSSSPQPKKKPL 309-323
4 SHLKSKKGQSTSRHK (367-381)	**4** RHKKLMFKTEGPDSD 379-393
without smoothing	
1 WFTEDPGPDEAPRMP (53-67)	**1** GPDEAPRMPEAAPPV 59-73
2 TEEENLRKKGEPHHE (284-298)	**2** KKGEPHHELPPGSTK 291-305
3 PNNTSSSPQPKKKPL (309-323)	**3** PNNTSSSPQPKKKPL 309-323
4 STSRHKKLMFKTEGP (376-390)	**4** RHKKLMFKTEGPDSD 379-393

Analogous studies were performed for the Znf 217 protein (plot not shown). Comparison of the epitopic peptides selected using the solvent accessibility algorithm with the two fragments achieved after surface probability calculations is shown in (Table 3.2.6). Regardless of peptide length, both epitopes described by Pavlik were present within the resulting ASS peptide file. The epitope sequence of the peptide No.1 was shortened by three residues in the case of generation of 15mers. The last three amino acids from the peptide No.2 were not incorporated into the peptide because these are the last C-terminal residues and their surface probability values could not be calculated by the algorithm, as discussed previously for C-terminal PDZ peptides. Predictions of surface probabilities for epitopic sequences form the p53 protein seemed to be less satisfactory comparing with the results achieved after calculations of solvent accessibility. However, the ASS predictions for the Znf 217 protein were comparable for both algorithms. In addition to the studies described above, more model proteins, containing known binding motifs have to be analyzed with both algorithms in order to compare and judge their performances.

Table 3.2.6: Comparison of peptides achieved from the Znf 217 sequence after calculations with surface probability (ASS) and solvent accessibility (Pavlik *et al.,* 2003) algorithms. Several different selection parameters applied for the ASS calculations are shown. Red: epitopic residues, which were predicted with the ASS, blue: epitopic residues, which were not predicted with the ASS.

Znf 217 protein	
Surface probability algorithm (ASS)	**Solvent accessibility algorithm**
pep 25 aa, peek 2.0	
smoothing 2.0	
1 TSVSPAPDKTKRPETKLKPLPVAPS (845-869)	PDKTKRPETKLKPLP 851-865
2 HLSNSMAQKRNYENFIGNAHYRPND (1021-1045)	NFIGNAHYRPNDKKT 1034-1048
without smoothing	
1 TSVSPAPDKTKRPETKLKPLPVAPS (845-869)	PDKTKRPETKLKPLP 851-865
2 HLSNSMAQKRNYENFIGNAHYRPND (1021-1045)	NFIGNAHYRPNDKKT 1034-1048
pep 15 aa, peek 2.0	
smoothing 2.0	
1 SPAPDKTKRPETKLK (848-862)	PDKTKRPETKLKPLP 851-865
2 NYENFIGNAHYRPND (1031-1045)	NFIGNAHYRPNDKKT 1034-1048
without smoothing	
1 SPAPDKTKRPETKLK (848-862)	PDKTKRPETKLKPLP 851-865
2 NYENFIGNAHYRPND (1031-1045)	NFIGNAHYRPNDKKT 1034-1048

3.2.5 Summary

The analysis of 25 protein sequences containing continuous binding motifs for different protein modules revealed that their majority was successfully predicted as surface probable with the ASS software tool. Since none from the existing algorithms is perfect enough to predict all possible, surface exposed regions in a protein, the achieved results are fully acceptable and satisfactory. In addition, the option to apply various parameters allows for more or less stringent calculations. The choice of optimized parameters is a compromise between a total number of peptides that have to be chemically synthesized and accuracy of predictions observed for different thresholds applied in the model systems. The peak height thresholds will eliminate motifs having lower surface probability values than a defined limit. High thresholds will decrease the amounts of peptides to be synthesized and will allow for generation of peptides only from highly exposed regions. On the other hand, application of very restrictive parameters will reduce chances for generation of those potentially bio-active motifs, which are less surface exposed. Such motifs can be selected by reduction of peak height threshold and generation of longer peptides, if desirable. Nevertheless, this will increase the total amounts of peptides in the resulting file.

A feature of the ASS software, which definitely needs to be improved, is its inability to predict surface probabilities for the N- and C-terminal protein sequences. To calculate the value for a certain amino acid (n) the algorithm involves values of surrounding amino acids (n-2 and n+3). Such calculations are obviously impossible for the first two and the last three residues, therefore their value is equal 0 (Emini *et al.,* 1985). This problem was observed for binding motifs of PDZ modules, as they usually contain the last C-terminal amino acids. Even in the case of the ErbB-4 protein described earlier, where the C-terminal region was scored as extremely surface-probable, the last three residues that are important for binding to PDZ domains were not incorporated into the generated fragment. As a proposal to overcome these problems, the N- and C-termini should be automatically included in the peptide output file.

Comparison of the ASS calculations with protein structures revealed that several regions displayed on a protein surface, especially loops connecting other structural motifs, were correctly scored by the algorithm. Selections of loop regions increase chances to generate peptides with potentially functional, linear epitopes, which are expected to predominantly occur within these regions. Nevertheless, to rely on the comparisons of protein structures with computational surface probability predictions could be questionable, as the structural data are frequently obtained for proteins bound with their ligands. The effects of structural changes occurring upon complex formation (Gohlke *et al.,* 2004) are unpredictable using the surface probability algorithm and could lead to inappropriate interpretation of results. Most

importantly, it was revealed that certain structural motifs like α-helical and β-strands were not correctly qualified as surface-probable, despite their evident exposure deduced from a protein structure. It could be hypothesized that difficulties with predictions for α-helices could be related to their “dual character”, defined by alternating, buried and exposed amino-acid residues. Despite certain limitations, the majority of results achieved using the ASS software was satisfactory. The ASS can be proposed as a useful tool, helpful in genome spanning approaches toward selections of novel binding ligands. Such application will be supported by high-throughput performance of the ASS, as hundreds of protein sequences can be analyzed simultaneously.

3.3 High-throughput affinity selection tests of the phage displayed protein library on the SPOT peptide arrays

The affinity selection experiments described in chapter 3.1 were performed using the model phages and the four peptide-ligands only. In order to test the phage enrichment in a more complex and realistic situation, a high-throughput panning experiment was carried out, employing the entire human brain T7Select10-3b phage library and screening against a peptide array containing a total of 400 spots. The primary information to be acquired from this experiment was the performance of the developed panning process as judges by the quantities and types of phages eluted from each single peptide-spot. It was suspected that those peptides which are recognized by expressed protein domains in the phage library will trap a higher number of phages and that this feature may be exploited to select only appropriate spots for further analysis. Additionally, an effort was taken to answer the question whether the analysis by PCR, AGE and clone sequencing is practical for the reliable identification of specifically bound phages enriched after only one round of affinity enrichment in this high-throughput situation.

In the first step, the peptide arrays were assembled on cellulose membranes as described in section 2.2.6. A set of sequences derived from naturally occurring human proteins expressed in brain tissue was used as the source of peptides. A list of 533 proteins that have been previously identified from mouse brain by 2D gel electrophoresis and mass-spectrometry technologies was provided by the collaborating partner (Klose *et al*, Charetie Universität Klinikum, Berlin, manuscript in preparation). Unknown or hypothetical proteins as well as proteins obviously involved in signalling or cell motility events were preferred. From the entire list of the mouse brain proteins, 119 were chosen for the search of their human homologues by Blast search. From the resulting list of human proteins, 98 sequences were chosen as candidates for peptide generation. Additionally, 14 sequences of the large size human brain KIAA proteins were chosen from the *H.U.G.E* database (http://www.kazusa.or.jp/huge/). The complete list of the Gene Bank database accession numbers (gi) is shown in Appendix 2. The Accessible Surface Scanner software was then applied to the 112 protein sequences to predict the exposed protein regions that were subsequently utilized for generation of peptides. The scanning parameters (peak height 2.0, minimum width 1.0, smoothing 2.0 and peptide length 25 aa) were chosen as described in the previous section. 1200 peptides were acquired, which were chemically synthesized on three cellulose membranes. Each membrane contained 400 different peptides in 16 lines of 25 spots each plus 25 replicates of a control peptide (GTPPPPYTVG), which is a ligand for the WW/YAP module. Membrane No.1 (see Appendix 3) was used for the test panning with the human brain T7Select10-3b

phage library containing 1.2 x 10^7 pfu/ml of the primary recombinants, which was previously amplified to give higher phage representation (1.1 x 10^{11} pfu/ml). The other two membranes were preserved for future investigations. The blocking and washing of the membrane and phages, as well as the detailed panning procedure are described in Section 2.2.4.4. After panning was completed, the spots were cut out and placed separately in the wells of 96 deep well blocks filled with TBST. In order to preserve the order of spots as it is on the cellulose membrane and allow an easy tracking of the peptide sequences, the spots were successively cut out, starting from the upper left corner of the membrane (spot No.1) and placed in the appropriate position of the deep well block (blocks PL1-PL5). Thus, spot No. 97 was then placed in the position A1 of the PL2 block (PL2A1). The bound phages were simultaneously eluted from the peptide-spots with 1 % SDS solution. In order to determine the total numbers of the phages eluted from each peptide-spot, series of dilutions (up to 1:10^4) were prepared and phages were poured out on the Petri agar plates with bacterial lawn, as described in 2.2.4.3. Phage titers found ranged from 10^4 pfu to 10^7 pfu and were evidently much higher than those observed in the previous experiments in which only four model peptide spots and a much less diverse phage mixture were used (see 3.1.2). The comparison of phage quantities recovered from the four control peptides in this experiment, with those determined after panning of model phages diluted at ratios 1:10^2 and 1:10^4 on the same peptide, showed an increase of two and three orders of magnitude for peptide-spots in this experiment (Table 3.3.1). Complete set of data showing amounts of phages eluted from each peptide from the membrane No.1 are shown in Appendix 3.

Table 3.3.1: Comparison of amounts of phages eluted from the control peptide-spot in the high-throughput panning and two previously performed model pannings, where the three model phages were diluted 1:10^2 and 1:10^4 with the Fas, IL-4 and Ezrin phages

Peptide	**Total quantities of eluted phages (pfu/spot)**		
	High-throughput panning	1:10^2 (Table 3.1.3)	1:10^4 (Table 3.1.3)
GTPPPPYTVG	2.19x10^6	4.9x10^4	6.5x10^3
	1.8x10^6		
	1.31x10^6		
	1.26x10^6		

The surprisingly high page titers obtained suggested that the washing steps and volumes of buffers applied were, in this case, inefficient to remove the unspecific bound background phages. This is concluded from the calculation of the total volume of buffer (30 ml) used in each step during washing of the membrane of which only approximately 75 µl could be

addressed to a single spot, whereas the volume used for a single spot wash in the former experiments was 250 µl. To investigate whether the volume of washing buffer affects the total number of recovered phages, four peptide spots from the membrane No.1 were reused for another panning experiment applying the same phage library. The phage panning was performed for each spot separately in the 1.5 ml Eppendorf tubes, using 250 µl of blocking and washing solutions. It was shown that the overall number of recovered phages could be reduced up to one order of magnitude in all four cases, but was still quite high ($5.8x10^5$ - $1.1x10^6$) see Table 3.3.2.

Table 3.3.2: Comparison of the total amounts of phages eluted from peptide-spots washed with different buffer volumes.

Peptide	**Spot position in 96 well plate**	**Total quantities of eluted phages**	
		Membrane (75µl/spot)	Single spots (250µl/spot)
GTPPPPYTVG	PL5B6	$1.8x10^6$	$5.8x10^5$
GTPPPPYTVG	PL5B7	$1.3x10^6$	$6.3x10^5$
RSFPDTVYKKLVQREKTLKVRGVDR	PL3G8	$7.0x10^6$	$9.0x10^5$
MKTRRTLKGHGNKVLCMDWCKDKRR	PL3A6	$1.3x10^7$	$1.1x10^6$

An additional panning experiment was carried out using a freshly prepared phage library, whose titer was reduced (1.5×10^{10} pfu/ml), comparing to that used for the previously described experiment. A fresh segment of the membrane No.2, containing the array of 28 peptide-spots was utilized. Except for the larger volumes of blocking and washing buffers (500 µl/spot), the panning procedure was carried out exactly as described for the membrane No.1. The quantities of the peptide-spot eluted phages were not as high as those observed for the membrane No.1 (10^6 or 10^7) and ranged from 10^4 for control peptides to 10^5 for the remaining 24 peptides. Peptide sequences, mother protein database identification code, as well as amounts of phages eluted from each peptide-spot are shown in Table 3.3.3. It is concluded that these are typical amounts of phages that can be recovered from the peptide spots after panning with the entire library, even when the combination of large volumes of buffers and less amplified libraries are utilized.

Table 3.3.3: Amounts of phages eluted from each of 28 peptide-spots washed with increased volume of buffer (500 μl/spot).

Protein ID	Peptide sequence	Amounts of eluted phages
gi3386485	APPTASPGPERKPPPASAPAPQPAP	1.33×10^5
	VVRSFRPDFVLIRQHAFGMAENEDF	1.09×10^5
	TLGGEKFPLIEQTYYPNHKEMLTLP	1.11×10^5
	SGNWKTNTGSAMLEQIAMSDRYKLW	2.98×10^5
gi12654186	KLKKPEAVRWEKLEGQGDVPTPKQF	5.62×10^5
	PTPKQFVADVKNLYPSSSPYTRNWD	2.47×10^5
	SSSPYTRNWDKLVGEIKEEEKNEKL	1.33×10^5
gi14424792	VSGRRRLLTVLLQAQKWPFQPSRDM	4.12×10^5
	SPSQLRKGPPVPPPPKHTPSKEVKQ	3.58×10^5
	VTSPVKAPTPSGQSIPWDLWEPTES	1.02×10^5
	QNPEEQDEGWLMGVKESDWNQHKEL	2.13×10^5
gi2967518	MSYQGKKNIPRITSDRLLIKGGKIV	3.4×10^5
	YDGQEQDELSFKAGDELTKLGEEDE	5.25×10^4
	TLNKDATKAATAAADFTAKVWDAVS	7.35×10^4
	VDFTQDSNYLLTGGQDKLLRIYDLN	5.55×10^4
gi16356636	DKLLRIYDLNKPEAEPKEISGHTSG	4.05×10^4
	TTLKDEGKLLEAKGPVTDVAYSHDG	4.65×10^4
	DGYSENNVFYGHHAKIVCLAWSPDN	6.6×10^4
	WSPDNEHFASGGMDMMVYVWTLSDP	1.59×10^5
gi12803340	TLSDPETRVKIQDAHRLHHVSSLAW	1.62×10^5
gi23221	KVRVIDPRKQEIVAEKEKAHEGARP	1.96×10^5
	FSRMSERQLALWNPKNMQEPIALHE	1.75×10^5
	NMQEPIALHEMDTSNGVLLPFYDPD	2.05×10^5
gi5701730	FEITDESPYVHYLNTFSSKEPQRGM	4.27×10^5
control	GTPPPPYTVG	4.65×10^4
	GTPPPPYTVG	5.85×10^4
	GTPPPPYTVG	8.7×10^4
	GTPPPPYTVG	1.41×10^5

The entire phage population eluted from a single spot will contain specific as well as background phages. The next goal of this experiment was the examination whether specific binders can be obtained after only one panning round, and can be identified among the number of analyzed clones using the PCR-AGE-DNA sequencing methods. For this purpose twenty phage plaques obtained from each of the two peptide spots PL3A6 and PL3G8 with the highest quantities of eluted phages, 1.3×10^7 and 7.0×10^6 respectively, were chosen for

further analysis. Of these, clones of the same or very similar sizes of the DNA inserts, which were represented more than once within the analyzed phage pool (see colour positions, Fig 3.3.1) were expected to be specifically enriched peptide binders and were selected for DNA sequencing.

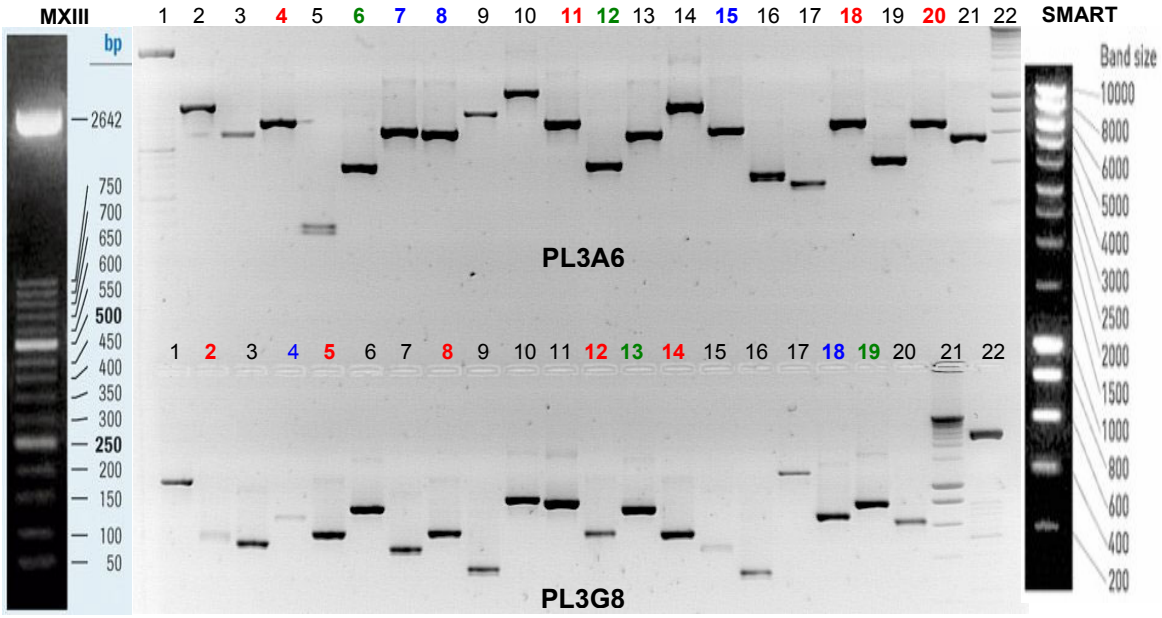


Fig 3.3.1: Ethidium bromide stained 1 % AGE analysis of DNA inserts of the phage populations recovered from the PL3A6 and PL3G8 peptide-spots. Coloured numbers indicate the phage clones selected for the DNA sequencing. The same colour indicates inserts showing the same or very similar sizes. **PL3A6)** In line 1: MXIII; 2-21: PCR fragments of 20 chosen phage clones; 22: SMART. **PL3G8)** In line 1-20: PCR fragments of 20 chosen phage clones; 21: SMART; 22: MXIII. MXIII: DNA marker (Boehringer Mannheim), SMART: DNA marker (Eurogentec).

The sequence analysis revealed that despite large DNA sizes, none of these sequences code for peptides or proteins larger than 65 aa, due to an early stop codon occurrence. Moreover single point mutations within phage capsid protein 10b were discovered in three cases. The Blast search showed that the majority of inserts represented fragments of non characterized genomic origin. Additionally, none of the inserts of the same or similar size was shown to code for the same protein or even peptide fragment. These results revealed that large numbers of irrelevant phages were present within the amplified library used for the panning experiment (Table 3.3.4).

Table 3.3.4: N-terminal fragments of translated sequences that are displayed on the phages eluted from PL3A6 and PL3G8 peptide-spots, whose DNA insert were subjected for sequencing. The sequence GDPNS represents C-terminus of the T7 phage coat protein 10b. Blue bold: single mutations; Red bold: the last amino-acid (**S**) of the phage sequence; Asterix: indicates termination codon; N.S.S: no significant similarity.

Fragment of translated aa sequence	Plate position /phage clone	BLAST result
G**N**PN**S**SL*VFS	PL3 A6/4	N.S.S
GDPN**S**SLNSSGYTNI**Y*N**	PL3 A6/6	emb|N59875
GDPN*AEKLSTL*EHDTTNFIFYHFLE	PL3 A6/7	emb|BC038857
GDPN**S**TNSLKT*KLH	PL3 A6/8	emb|BD130959
GDPN**S**NRESGSSSVPLSRALLQQHPVTVSCQNISVCHGCCP GFSHDNYHCKQKK*	PL3 A6/11	emb|CAJ04699
GDPN*AELCLLLFNFFFFF	PL3 A6/12	N.S.S
GDP**I**KPKRTPTKRQL*QKYAL	PL3 A6/15	emb|AX069004
GDPN**S**AEMCTQKVQRQxxxPSEAxxPQMRFGLEKLEKGMQD EL*T	PL3 A6/19	emb|AC012254
GDPN**S**ALQVRVFLLIFLVEGREEGEEQGRRV*E	PL3 A6/20	emb|AC084170
GDPN**S**QDLSIFYNRNVHHELDTSVCPRCCH*	PL3 G8/2	emb|CB117181
GDPN*AVWRGPQFQQSRCSHWPQSRHTCQGLWSRLRL*R	PL3 G8/4	emb|BQ440041
GDPN**S**A*GGLELP	PL3 G8/5	emb|AC024700
GDPN**S**AREGWRQTSPVVSKGAMCLIYKVS*	PL3 G8/8	emb|AL138108
GDPN**S**SPLVR*VTCIFYKGLGADLDKS*	PL3 G8/12	N.S.S
GDPN**S**EKAAPKMSTCSPIPEHKLSILGSGFRNHLCGKT*	PL3 G8/13	emb|AC141288
GDPN**S**ARLNTPHFRNTHLLGNF*	PL3 G8/14	emb|AC003003
GDPN**Y**IVMKSDKEIVGTLLGFDDFVNMVLEDVTEFEITPEGRR ITKLDQILLNGNNITMLVPGGEGPEV*	PL3 G8/18	emb|BG716585
GDPN**S***DPGKDQHANQCQ*	PL3 G8/20	emb|AC069040

Based on this result, the quality of the commercial phage library was heavily questioned. Therefore, phages from the mother library lysate were poured on a bacterial lawn. Subsequently twenty phage plaques were randomly chosen and subjected to DNA sequencing. Because of pour sequencing quality of the 5' ends of the inserts in many cases, the identification of cloning sites and correct open reading frames, thus also identification of possible stop codons occurring early at the 5' end of the insert was very difficult. Fifteen sequences were subjected for DNA versus DNA, BLAST homology search. Five remaining clones failed the sequencing. It was revealed that, only three clones (3, 14, and 15), were recognized in the databases as sequences coding for known or hypothetical proteins (Table

3.3.5). The remaining sequences were either uncharacterized or represented fragments of chromosomal DNA. Therefore, it was suspected that the majority of analyzed inserts code for short non-functional peptide fragments due to possible early stop codon occurrence, similarly as described in the above studies. The abundance of phages coding short peptide fragments is probably even increasing during the library amplification step.

Table 3.3.5: Results of BLAST homology search performed for fifteen randomly chosen phages from mother library lysate.

Clone	Description of homologous sequence	Database ID
1	H.s cDNA clone UI-CF-EN1-aek-c-04-0-UI 3', mRNA sequence	emb|CB306201
2	H.s DNA sequence from clone RP4-590F24 on chromosome 1	emb|AC027074
3	Tripartite motif protein 2 (RING finger protein 86)	emb|BC011052
4	H.s DNA sequence from clone RP11-262D11 on chromosome X	emb|BX119917
7	H.s mRNA; EST DKFZp686C03165_r1	emb|BX473784
8	H.s mRNA; EST DKFZp686L13127_r1	emb|BX471489
9	H.s bone morphogenetic protein receptor, type II (serine/threonine kinase), mRNA (cDNA clone IMAGE:5261941)	emb|BC035097
11	H.s cDNA clone cs02c09 3', mRNA sequence	emb|N48055
13	H.s chromosome 14 DNA sequence BAC R-671J11 of library RPCI-11 from chromosome 14 (Human)	emb|AL445363
14	H.s septin 3 isoform C mRNA, complete cds	emb|AF285108
15	H.s mRNA for KIAA1219 protein, partial cds	emb|BC047482
16	wn89d11.x1 NCI_CGAP_Ut1 H.s cDNA clone IMAGE:2453013 3', mRNA sequence	emb|AI922199
17	Sequence 20 from Patent WO9951727.	emb|AX015361
19	H.s 3 BAC RP11-705F7	emb|AC130003
20	UI-CF-EN1-adg-c-03-0-UI.s1 UI-CF-EN1 H.s cDNA clone	emb|CB851025

The phage panning, utilizing the entire T7Select10-3b phage library and larger peptide array, revealed a much higher complexity than the model pannings described in 3.1., where the mixtures of six phages and four peptide-spots only were applied. The elevated amounts of phages eluted from peptide-spots observed in the high-throughput panning were obviously the result of a higher diversity of this library, which provided more candidates for the specific as well as unspecific interactions. The broad range of phage titers (10^4 pfu - 10^7 pfu) obtained from the individual peptide-spots showed various preferences of different peptides for interaction with phage displayed proteins and peptides. This was not as noticeable in the earlier performed model experiments. An additional factor increasing the amounts of

recovered phages was certainly the inappropriate panning procedure. The increase of buffer volumes applied for the panning process (from 75 µl/spot to 500 µl/spot) led to positive effects on the elimination of unspecific binders and decreased the phage titers to at least one order of magnitude. Nevertheless, the total quantities of recovered phages were still much higher than that observed for the model system. Another factor relevant for the affinity selection experiments is the quality of the utilized library. The sequencing analysis showed that a large fraction of the phage population in the T7Select10-3b library displayed short peptide fragments instead of larger functional proteins or protein modules which are the expected potential binding partners for the membrane spotted peptides. The same library was successfully used for the affinity selection approaches as a source of peptides interacting with protein modules immobilized on the solid support (Kurakine and Bredesen, 2002). Theoretically, only one-third of all cDNA inserts code for protein/peptide sequences from the proteome. The construction of the library should favour a correct orientation; two-thirds then can be considered as “random peptides” originating from frame shifts upon ligation and incorrect insert orientation. In reality, however, the fraction of the proteome in the library must be much less, due to priming from untranslated regions of mRNA (Kurakine and Bredesen, 2002). Additionally, the T7Select10-3b library was reported to contain even rare mRNA sequences, which are represented in it with high probability (Rosenberg *et al.,* 1996 refer also to Soares *et al.,* 1994).

Since the high-throughput panning over the whole spot-peptide array carried out in this section involves only one round of phage selection, a library amplification step preceding the panning process was necessary in order to increase the number of rarely represented candidates. Such amplification may additionally increase the background phage populations that display short irrelevant peptidic sequences, because these may grow faster than other phages displaying proteins of larger size and cause phage assembly more difficult. The data obtained from the sequence analysis of phage clones eluted from the peptide-spots confirm the high abundance of such background phages.

Classical phage panning on a defined immobilized target involves several rounds of enrichment/panning, each round monitored by an ELISA assay for an increase of target specific phages. A corresponding practical monitoring of phage eluates in a multiplexed situation would be helpful to select those peptide ligands that successfully enrich phages from the library. A simple monitoring based on phage titers would be very practical, however, because of the too high background found after one panning round, this easy measure was not found to be indicative of a successful enrichment. Consequently, the reliable identification of specifically enriched phages in a real high-throughput performance would require a new appropriate methodology.

3.4 Target-Assisted Iterative Screening (TAIS)

The results described in this chapter are concerned with the evaluation of the Target Assisted Iterative Screening (TAIS), an elegant method developed by Kurakin for mapping of protein-peptide interactions (Kurakin *et al.,* 2002, Kurakin *et al.,* 2004). TAIS combines phage display and expression library screening formats and was successfully applied to screen phage libraries of peptides over proteins/domains immobilized on a solid support (i.e. Sepharose beads). In the first step, a phage-displayed peptide-size cDNA library in solution is incubated with the immobilized target protein. Phages displaying peptides interacting with the target protein remain bound, whereas the non-bindings once are washed away during the bio-panning process. In the second step, the bound pre-selected phages are eluted and plated onto a bacterial lawn. Next, the plaques formed are lifted onto a nitrocellulose membrane by blotting the bacterial lawn with the membrane. In the meantime, the target protein in solution is biotinylated and complexed with a streptavidin conjugated reporter molecule such as i.e. alkaline phosphatase; the complex is used as a detection reagent for those plaques, which can bind the displayed peptide on the membrane. Subsequent analysis of DNA inserts of positive phage clones allows identification of the displayed peptide. The flowchart in Fig 3.4.1 is describing principle of this method.

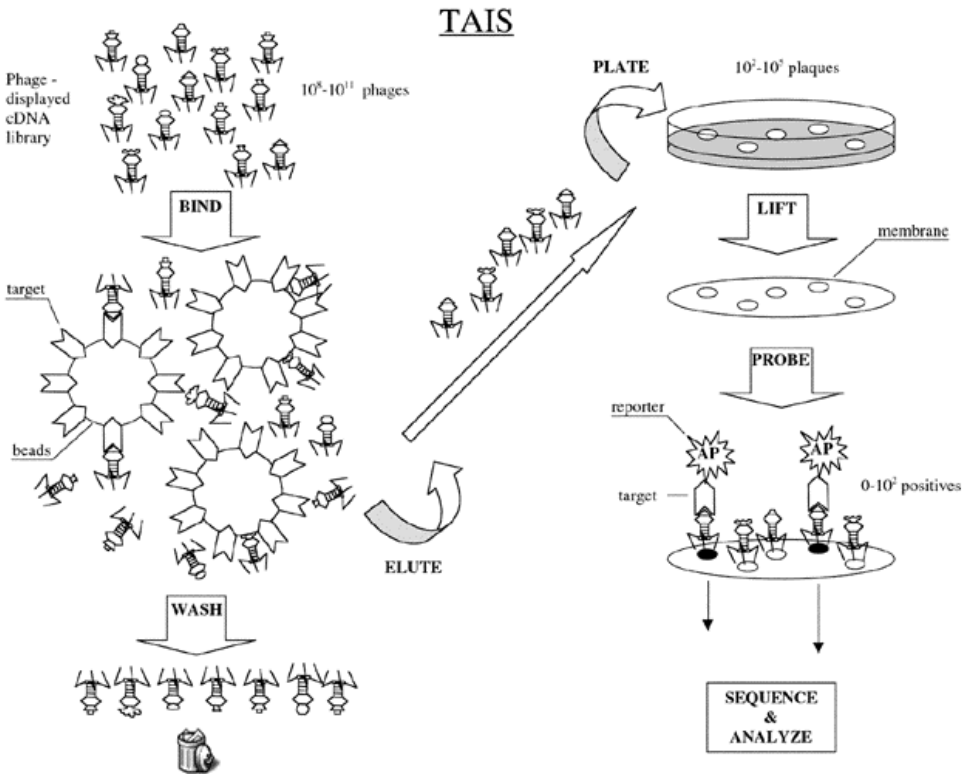


Fig 3.4.1: Schematic flowchart explaining the Target Assisted Iterative Screening process (Kurakin *et al.,* 2002).

Here, a reversed mode of the TAIS technology was applied for screenings of phage displayed repertoires of protein domains over immobilized peptide-ligands. For this purpose, two sets of peptides having the same amino-acid sequences were chemically synthesized on different types of modified cellulose membranes. One set contained peptides that remain covalently attached to the solid support for phage affinity selections in the panning step. The second set contained N-terminally biotinylated peptides, which were cleaved from the membrane through formation of a C-terminal diketopiperazine and released into solution phase (see 2.2.6). In order to test the performance of this novel system in combination with target peptide arrays, a panning was carried out over three 10-15 aa long peptides that are ligands for the three model domains (WW/YAP, WW/FE65, EVH1) (Table 1.3.1 in section 1.3). The model phages were supplemented to the T7 human brain phage display library at a ratio of $1:10^6$.

Panning was performed as described in 2.2.4.4. Phages enriched in the first selection step were separately eluted from peptide-spots using 100 μl of 1 M guanidine buffer. Serial dilutions of the eluate were prepared immediately. Aliquot of 100 μl from dilution $1:10^3$ were plated on a host bacterial lawn on large Petri dishes (diameter 15 cm). Despite high phage dilution applied, large amounts of plaques were obtained on each plate, often fused with each other. The phage titers were estimated to be 10^5 pfu / 100 μl of the eluate. Surprisingly, this is at least one order of magnitude higher than titers (10^3-10^4) achieved for the same peptides in similar panning experiments described in section 3.1.2 (Table 3.1.3). However it is comparable to titers (10^4-10^5) obtained for the peptide-ligand of the WW/YAP domain, after panning on the cellulose membrane containing 28 spots (see Table 3.3.3 in section 3.3). The observed variations could be explained by the fact that all of mentioned experiments were carried out using different batches of phage lysates. It is possible that the lysate used for the above panning contained higher concentration of background phages which are responsible for higher phage titer (see discussion in section 3.3).

In the second step, plaques were transferred onto nitrocellulose membranes. Biotinylated peptides in solution were complexed with streptavidin that is conjugated to alkaline phosphatase reporter (STR-AP), as described in 2.2.4.6. Theoretically, each streptavidin molecule binds four molecules of peptide, as it contains four biotin binding sites. Resulting peptide tetramers were incubated with the pre-selected phages blotted on the nitrocellulose membranes. After addition of the alkaline phosphatase substrate (BCIP/MMT), distinct color signals were expected for peptide specific phage plaques. In the cases of the EVH1 and WW/FE65 peptide-ligands, no signals indicating specific interactions of these peptides with corresponding protein domains on the phages were observed. Six well visible signals were found on the membrane with plaques obtained from panning on the WW/YAP peptide (Fig 3.4.2).

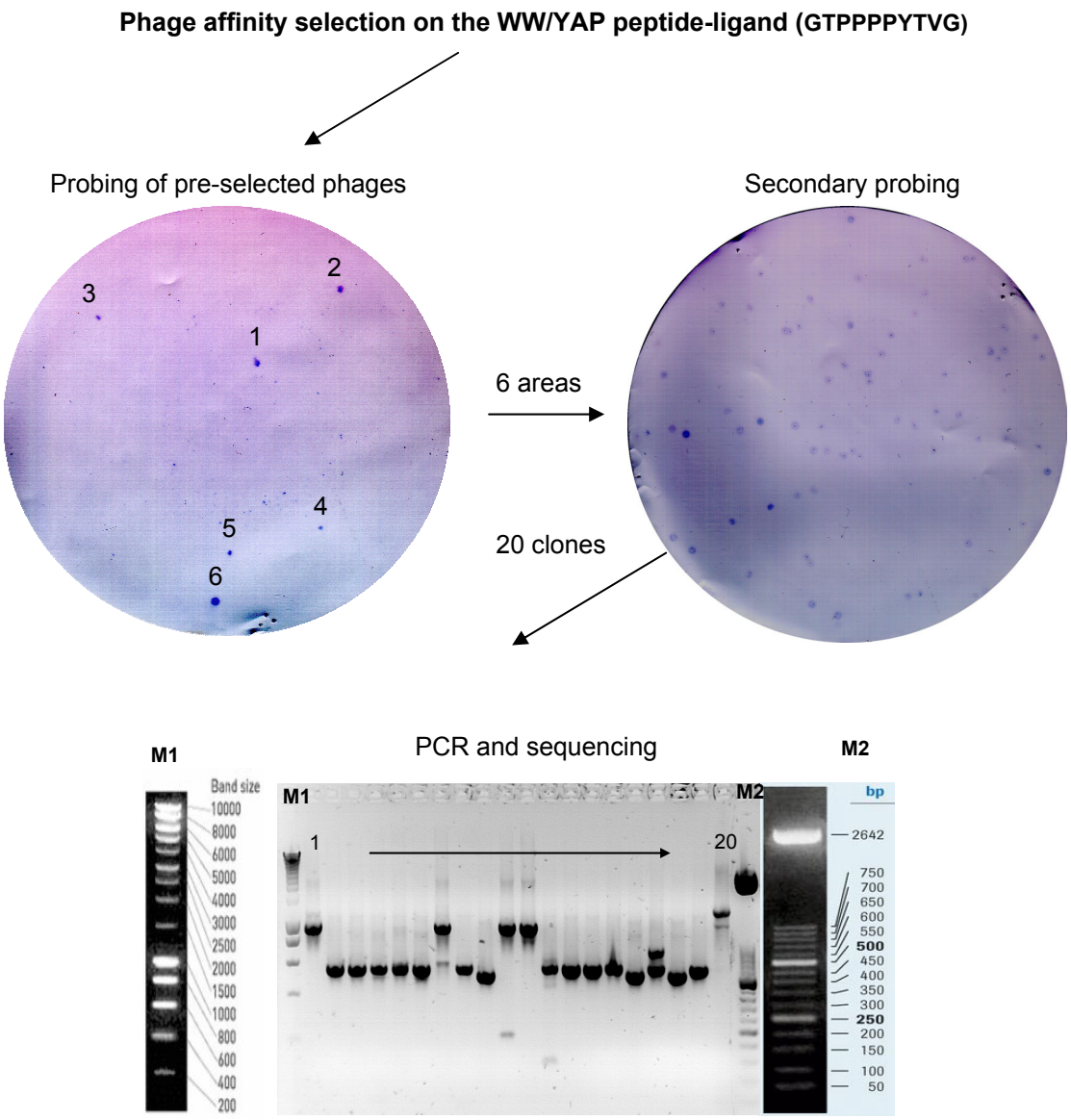


Fig 3.4.2: Schematic view of the entire TAIS process performed using the WW/YAP peptide as a ligand. M1) DNA marker SMART (Eurogentec). M2) DNA marker MXIII (Boehringer Mannheim).

Because of difficulties in isolation of single positive plaques, due to their overgrowing with other phages, square areas of top-agarose containing the plaques of interest as well as surrounding plaques, were cut out from the plate and collected in the same tube. Phages were extracted using PEB buffer (see Table 2.1.10.1). After experimentally determining a suitable dilution ratio of $1:10^6$, phages were poured again on a plate with the bacterial lawn, in order to achieve sufficient plaque separation. Plaque transfer onto a nitrocellulose membrane and probing with the peptide-STR-AP tetramer was repeated as described above and resulted in identification of a large number of positive plaques (Fig 3.4.2). Twenty of these plaques were picked from the agar plate using sterile tooth sticks and placed in separate tubes. Phages were eluted from agar pieces and their DNA was extracted, as

described in 2.2.3.5. Phage DNAs were used as templates for PCR followed by insert sizing by AGE. Subsequently all inserts were chosen for DNA sequencing and sixteen sequences were used for BLAST homology search of gene databases. Unexpectedly, homology search revealed that none of the selected clones contained DNA encoding the WW domain of the YAP protein that was supplemented to the pre-panning phage population. Three verified inserts contained DNA sequence coding for the Itchy homolog E3 ubiquitin protein ligase and the remaining thirteen inserts code for the Nedd 4 like ubiquitin protein ligase (Table 3.4.1). Both proteins belong to the same family of protein ligases, which contain four WW domains (class I), which bind the (L/P)PxY (x-any amino-acid) consensus motif.

Table 3.4.1: Results of BLAST search performed for DNA sequences of selected phage clones.

WW domains	Clone	Protein name	Swiss prot	Synonyms
1, 2, 3	1	Nedd-4-like ubiquitin-protein ligase	Q9H0M0	WWP1[a)], AIP5[b)]
1, 2	2	Nedd-4-like ubiquitin-protein ligase	Q9H0M0	
1, 2	3	Nedd-4-like ubiquitin-protein ligase	Q9H0M0	
1, 2	5	Itchy homolog E3 ubiquitin protein ligase	Q96J02	AIP4[c)] , NAPP1[d)]
1, 2, 3	6	Nedd-4-like ubiquitin-protein ligase	Q9H0M0	
1, 2	7	Nedd-4-like ubiquitin-protein ligase	Q9H0M0	
1, 2	9	Nedd-4-like ubiquitin-protein ligase	Q9H0M0	
1, 2	10	Nedd-4-like ubiquitin-protein ligase	Q9H0M0	
1, 2	11	Nedd-4-like ubiquitin-protein ligase	Q9H0M0	
1, 2	12	Nedd-4-like ubiquitin-protein ligase	Q9H0M0	
1, 2, 3	13	Nedd-4-like ubiquitin-protein ligase	Q9H0M0	
1, 2	14	Nedd-4-like ubiquitin-protein ligase	Q9H0M0	
1, 2	15	Nedd-4-like ubiquitin-protein ligase	Q9H0M0	
1, 2	16	Itchy homolog E3 ubiquitin protein ligase	Q96J02	
1, 2	18	Itchy homolog E3 ubiquitin protein ligase	Q96J02	
1, 2	19	Nedd-4-like ubiquitin-protein ligase	Q9H0M0	

[a)] WW domain-containing protein 1; [b)] Atropin-1 interacting protein 5; [c)] Atrophin-1-interacting protein 4; [d)] NFE2-associated polypeptide 1

The positive phages identified here, display proteins that contain at least two WW domains as deduced from the analysis of DNA inserts. A very efficient enrichment of these multi WW-domain proteins, displayed in multiple copies on the phage coat, was most probable because of the high avidity of binding to several peptides on the spot. In contrast to this, the avidity of a single WW domain of the YAP protein should be respectively lower. Moreover, it is possible that phages presenting both identified protein ligases are very abundant within the

T7 brain library. High avidity of these phages combined with a high representation within the screened population, could result in strong competition with the WW/YAP domain for binding to the peptide-ligand. However, the amount of spot localized peptide (~5 nmol) is sufficient to select large number of different binders. It is therefore more realistic that the YAP phages were strongly under-represented within the post-panning population and could be identified only if larger number of phage clones were analyzed.

Analogously, screening of the entire post-panning phage populations obtained from peptide-ligands for the EVH1 and WW/FE65 domains, may also allow the identification of these model proteins. This would however, drastically increase the number of membranes and amounts of reagents that would be necessary to perform such a high-throughput scale screening. The average number of separate phage plaques (about 1 mm in size), which can be evenly distributed on one large Petri dish, is approximately 1000. Theoretically, a screening of the entire post-panning population from one peptide containing $1x10^5$ pfu, would then require about 100 plates and membranes. Costs of such approach would definitely be too high. Optionally, the two step process as shown in Fig 3.4.2 could be applied, plating all phages from the panning (up to 5000 pfu per plate) and selecting individual phages only from the secondary plates. Such a procedure can reduce numbers of plates and membranes to about 40, according to the following calculations (first screening: $1x10^5$ pfu: 20 plates with 5000 pfu each, second screening again 20 membranes).

The TAIS technology was considered here to be used in combination with high-throughput panning approach utilizing membranes containing large number of 25mer peptides, as described in section 3.3. In addition to the model experiments described above, several screenings of the T7 brain library were performed using 25 aa long peptides presented in Table 3.4.2, which were selected using the Accessible Surface Scanner software as described in chapter 3.2.

Table 3.4.2: Peptides used for TAIS experiments

No.	Peptide sequence	Source protein ID
1	VGNRFQTARFYRDVLGMKVLRHEEF	gi14250348
2	DLQKSLNYWCNLLGMKIYEKDEEKQ	
3	CPQKELPDLEDLMKRENQKILTPLV	
4	EGSKLLDDAMAADKSDEWFAKHNKP	
5	GPRSPRSASPGPASTGPRHPQVAEG	gi1504033
6	CHGRHEACEARGRRPRLLNPGRRGG	
7	QIAARTRRRGGGRMERKGSAAGAKG	
8	RKGSAAGAKGNPSPPAAGEGQRPPP	
9	AHEFKSQGAQCYKDKKFREAIGKYH	

No.	Peptide sequence	Source protein ID
10	ALLELKGLLPPPGERERDSRAASPA	
11	ERDSRAASPAGALKPGRLSEEQSKT	
12	VNYERVKEYCLKVLKKEGENFKALY	
13	HLGDYDKALYYLKEARTQQPTDTNV	
14	QPTDTNVIRYIQLTEMKLSRCSQRE	
15	NSRPGKNGNPVAENFRMEEVYLPDN	gi21755897
16	SVDPYMRCRMNEDTGTDYITPWQLS	
17	EESKHTNLTKGDFVTSFYWPWQTKV	
18	SFYWPWQTKVILDGNSLEKVDPQLV	
19	TSELGFDAAINYKKDNVAEQLRESC	
20	ISQMNENSHIILCGQISQYNKDVPY	
21	AIQKERNITRERFLVLNYKDKFEPG	

As in the cases described above, due to large amounts of post-panning phages, only an aliquot of each population (100 µl from a $1:10^3$ dilution) was subjected for probing with the corresponding soluble peptide-STR-AP tetramers. None of the resulting filters, however, showed clear signals for peptide interacting phages. In several cases, instead, atypical, “daunt shaped” signals were observed for all phage plaques of the plate transferred onto the nitrocellulose membrane (Fig 3.4.3). These signals were most probably reflecting an unspecific interaction of peptide tetramers with the phage and/or displayed proteins.

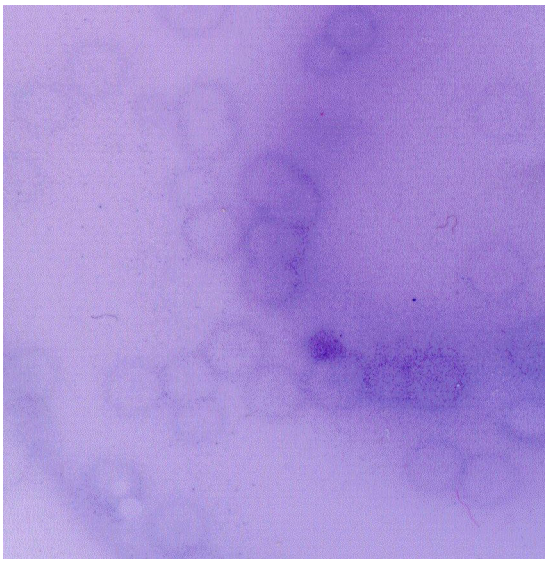


Fig 3.4.3: A section of nitrocellulose membrane showing phage plaques obtained from peptide No.20 and showing high unspecific interaction with soluble peptide-STR-AP tetramers.

Thus, additionally studies have to be conducted, attempting a reduction of unspecific phage background and a reduction of the total number of post-panning phages to be analyzed. This could be achieved by testing and application of more stringent bio-panning for cellulose spots or by application of other solid supports for chemical peptide synthesis, which could “trap” lower number of phages that are irrelevant for interaction with the target. Overcoming of this problem would significantly decrease costs of screenings. This would allow using TAIS as high-throughput method. Moreover, several steps of this process could be automated by application of new robotic technologies. Following high-throughput parallel affinity selection, separate phage elution and plating, the whole populations of post-panning plaques could be transferred into micro-plates for phage extraction, using colony picking robots. Arraying of phages on the nitrocellulose membranes would be carried out using automated printing pin tools (96 format or higher). Optionally, phages would be printed on agar plates with bacterial lawn and development plaques would be manually lifted onto the membranes. Plaques showing positive interaction signals would be isolated manually and subjected for DNA analysis. This concept could be applied after reduction of costs of TAIS and would allow identifying large number of interacting proteins simultaneously.

In conclusion, Target Assisted Iterative Screening is an efficient method, definitely showing a significant improvement over the type of crude analysis described in the previous section, where phage plaques were only randomly picked from the agar plate and subjected for PCR and sequence analysis, in order to identify enriched phages within post-panning populations. Although possible, complete screening of large post-panning phage populations using this technology would require enormous number of expensive membranes and reagents. Currently, this economical aspect rules out TAIS as a high-throughput, genome and proteome spanning method, which could be utilized in combination with cellulose peptide arrays. Other, rapid and cost effective methods must be developed in order to achieve this goal.

3.5 DNA microarray biochip based screening

3.5.1 The concept

The experiments described before as well as studies performed by Bialek (Bialek *et al.,* 2003) proved that peptides synthesized on cellulose membranes can be successfully utilized for affinity selection of phages displaying protein domains. The advantage of SPOTs is a high amount of peptides deposited on a small single spot (0.5-5 nmol), resulting in a high local peptide concentration and high avidity of peptide-protein binding that increases the chance for efficient enrichment of proteins which are less represented within a screened library. On the other hand, phage background is quite high due to the porous structure of the cellulose, most probably by trapping phages between its fibers. Performed investigations showed that, starting from phage populations containing approximately 10^{10} pfu, the number of phages eluted from a single peptide-spot after one round of affinity selection ranged from 10^4 to 10^6 pfu. Only one round of selection was preferred in order to reduce the phage competition process, which will otherwise result in a loss of weaker binders and overgrowing with the dominant populations as commonly observed during the classical consecutive panning procedure. Therefore, identification of specifically bound candidates within such diverse mixtures appeared to be very challenging. The standard procedures (PCR, sequencing) that are used for phage clone verification after multiple rounds of selection will require enormous number of clones to be analyzed in order to “sift” specific binders from background phages (see 3.2). Also, a novel protein-protein interaction screening method (TAIS) developed by Kurakin (Kurakin *et al*., 2002) has been shown to be economically unfavorable when large numbers of panning experiments were to be performed in parallel (see 3.4). Therefore, a high-throughput screening of the entire phage displayed library of protein domains versus high density peptide arrays, as anticipated in this work, requires very competent and efficient analytical tools for a rapid, automated and sensitive, simultaneous analysis of whole populations of phage displayed protein domains, enriched in only a one step affinity selection.

Post-panning phage populations are enriched for target (peptide)-specific binders (proteins/genes) and, thus, differ significantly from pre-panning populations. Similar differences concerning gene representation can be observed for cDNA pools obtained from two cell population that were exposed for different physical conditions or represent different physiological states (i.e. healthy and cancer). Variations in transcript level are usually studied using expression profiling method utilizing DNA microarray biochips. Microarray biochips containing thousands of genes are either commercially available or can be fabricated by printing of oligonucleotides or PCR products on a modified glass surface using

microarrayers. Expression profiling analysis involves isolation of mRNA from both test and reference cells and their separate labeling with different fluorescent dyes (i.e. Cy3-green and Cy5-red). Combined labeled cDNA samples are then hybridized to a set of DNA probes arrayed on a biochip and subsequently scanned, using laser scanner. Identification of over or under-represented genes is possible via computational measurement of relative fluorescence intensities emitted by the excited dyes built into the DNA structures (Fig 3.5.1). This attractive method was considered here as a powerful and potentially adequate tool that could be exploited for multiplexed simultaneous monitoring of pre- and post-panning phage representation and thus immediate parallel identification of enriched phages. Therefore, a primary goal of this work was the investigation whether this technology could be efficiently applied in the context of a high throughput phage panning approach.

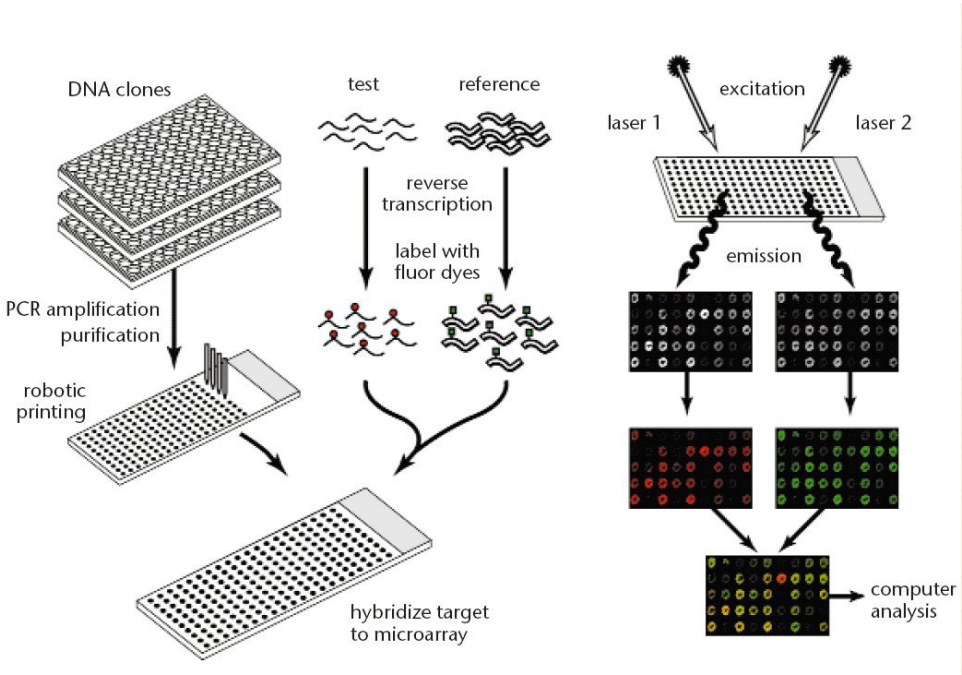


Fig 3.5.1: A schematic view showing expression profiling method

The experimental concept involves that after multiplex affinity enrichment, phage populations are separately eluted from each peptide-spot (post-panning) phage DNA will be isolated, labeled during PCR amplification with the cyanine dyes, Cy3 (green) or Cy5 (red) and mixed with an equal amount of correspondingly different labeled reference DNA obtained from the original phage library that was used for panning (pre-panning). This mixture will be hybridized with a DNA microarray biochip (Fig 3.5.2) .Compared to the pre-panning DNA, genes coding for peptide enriched proteins should be over-represented within the post-

panning DNA. The DNA sequences of the peptide-specific, over-represented genes and thus the protein identities are then easily obtained from the gene lists provided with the biochip, displaying all DNA probes specific for the ORFs coding for the proteins of interest.

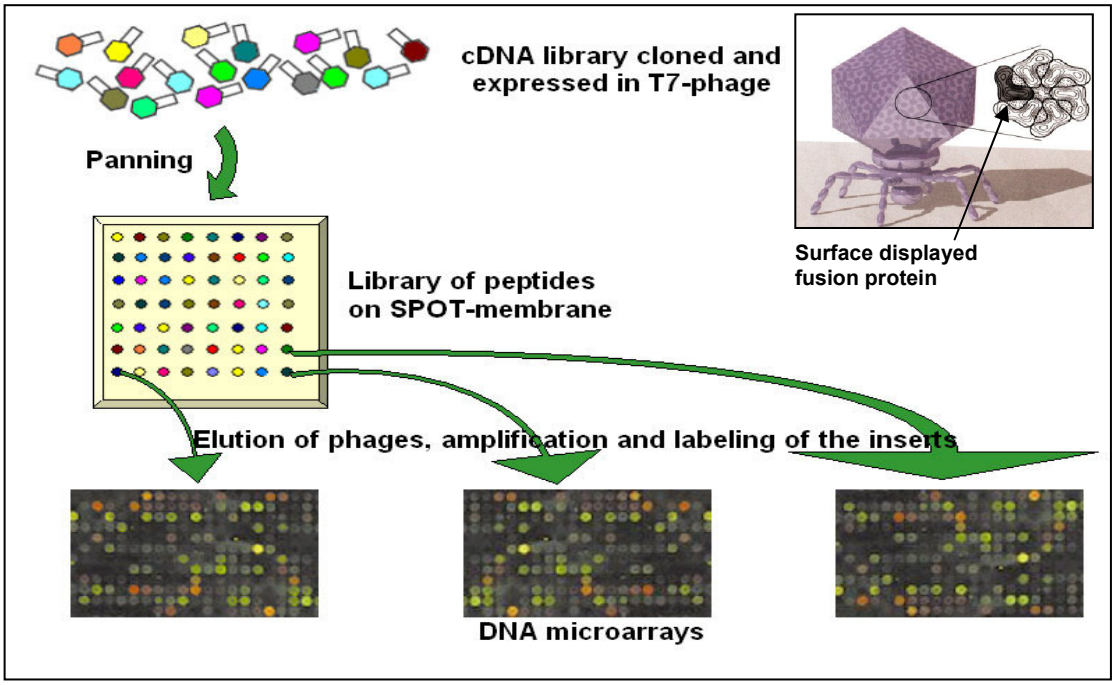


Fig 3.5.2: Schematic illustration of the experimental concept.

The application of this new methodology for genomic and proteomic approaches, requires highly representative DNA microarrays. Several companies, i.e. Affymetrix, offer commercial high-density oligonucleotide biochips. These however, can not be successfully applied for this screening approach (Bialek, GBF Braunschweig, unpublished data). The reason for this is a random priming strategy used for preparation of the cDNA inserts in the T7Select10-3b human brain library (Fig 3.5.3). The random primers anneal to various sites on the mRNA, providing cDNA fragments of different sizes and locations. The probe sequences, however, chosen for the chemical oligonucleotide synthesis are typically from the 3' end of the transcript, near to the poly-(A) tail (Lipshutz *et al.*, 1999; Lipshutz *et al.*, 1995). Use of the oligonucleotide arrays for the above approach therefore would bias the insert detection possibilities to those that are located on the 3' end of the transcript only. The only viable alternative to the oligonucleotide microarrays is an array containing full length cDNA sequences. Such biochips however, are not commercially available and must be self fabricated or purchased from custom services.

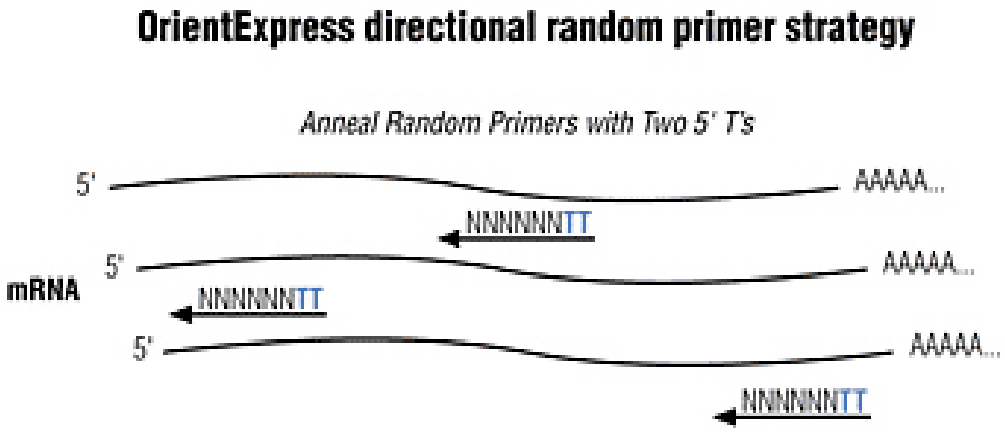


Fig 3.5.3: Schematic illustration of the OrientExpress random priming strategy used for preparation of the cDNA T7Select10-3b human brain library. Image was taken from Novagen's catalogue.

3.5.2 Proof of principle

An initial biochip based experiment was performed by Bialek (Bialek *et al.*, 2003), in order to test whether the expression profiling method could be exploited to detect more sensitively and in a highly parallel mode peptide-specific enrichment of phages. Simple microarray slides were prepared by contact printing of cDNAs that correspond to the three model domains (EVH1, WW/FE65, WW/YAP) plus a heavy chain of a mouse antibody as a control. The human brain T7Select10-3b library was doped at the ratio $1:10^4$ with phages displaying the two model domains EVH1 and WW/FE65. This was then panned over the peptide-spots that are ligands for these two domains. Peptide-spot bound phages were eluted and first propagated. DNA obtained from the pre-panning phage library and the phage populations eluted from each peptide ligand, were labeled with the Cy5 and Cy3 dyes, respectively. Additionally, a mouse control DNA was labeled with both Cy5 and Cy3 in two separate reactions. One micro-gram each of labeled DNAs from one of the peptide-specific post-panning phage inserts, of the pre-panning DNA and the two control DNA preparations were carefully combined and the mixture was applied to a microarray slide (Fig 3.5.4). Hybridization of the DNA mixture obtained from the peptide ligand for the EVH1 domain showed (pep-E) that the DNA probe on the slide corresponding to the EVH1 domain emitted a green fluorescent signal, indicating that more Cy3 than Cy5 labeled DNA hybridized to this probe. This means that DNA coding for the EVH1 domain is more abundant in the post-panning pool compared to the pre-panning pool. Thus, it clearly indicates enrichment of the EVH1 domain presenting phages on the expected target peptide. In contrast, the DNA probe corresponding to the WW/F65 domain emitted a red signal, indicating preferential hybridization with the Cy5-labeled, pre-panning WW/F65 DNA. This revealed the absence of

the WW/FE65 displaying phages within the post-panning population confirming the lack of its interaction with the pep-E ligand.

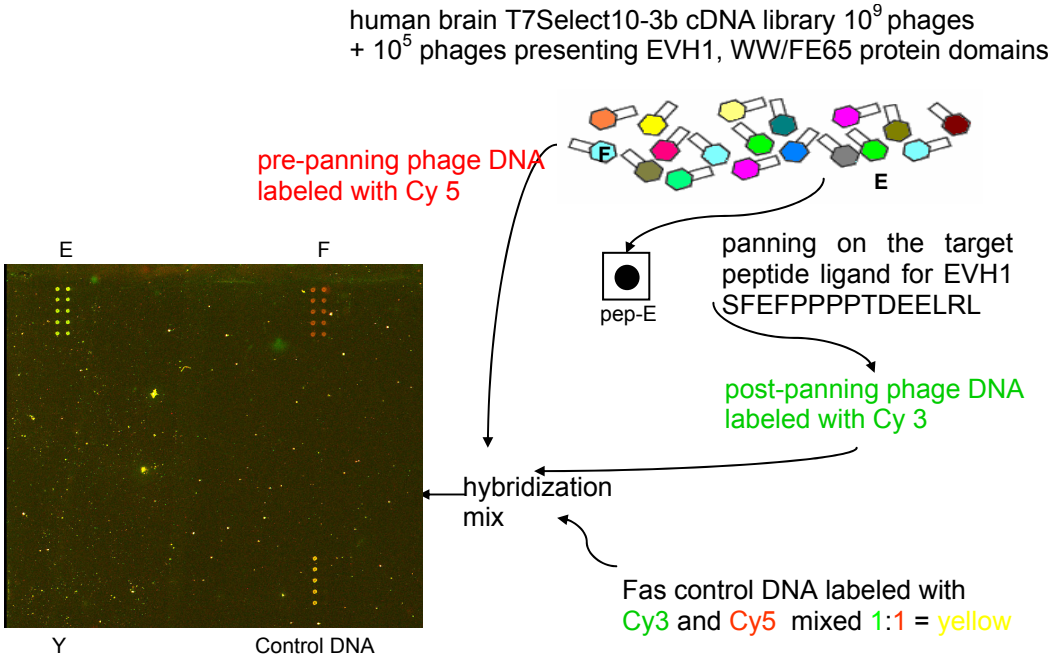


Fig 3.5.4: Schematic picture showing the panning process for the phage library supplemented with the model domain phages, on the peptide ligand for the EVH1 domain (pep-E) and results obtained from the differential hybridization of the pre- and post-panning phage library DNA to the DNA probes printed onto a glass slide. Short name code: EVH1 (E), WW/FE65 (F); WW/YAP (Y).

The control DNA hybridized to nearly equal amounts of Cy5- and Cy3-labeled DNA resulting in a yellow color confirming proper experimental performance. The DNA probe corresponding to the WW/YAP domain did not emit any signal, which is in agreement with the fact that the phages displaying this domain were not added into the library. It also suggests that this domain is not present in the library or only in such a low amount that it could not be detected. The panning of the WW/FE65 domain on its target peptide was also investigated by using DNA microarray analysis; however, the enrichment signal was much lower as compared to EVH1. This can be explained by weaker interaction of the WW/FE65 domain with its peptide, when compared to the interaction of EVH1 with its ligand (see previous chapters).

The phage propagation step included in this proof-of-principle experiment is quite problematic because it could increase number of background phages (see section 3.3). The attempts to omit the phage propagation step and to isolate and purify the DNA directly from the peptide-spot eluted populations using precipitation with isopropanol and commercial kits

(Millipore), however, gave non-satisfactory results (data not shown). In many cases, none or very weak signals indicating phage enrichment were achieved after hybridization with the probes on a biochip, suggesting an obvious problem with the direct purification of DNA. Most probably, the amounts of phage DNAs, isolated directly from relatively small amounts of eluted phages (10^5 pfu) were not enough to yield the amounts of labeled PCR products required for a sufficient fluorescent signal after the hybridization process. It is possible that large percentage of the phage DNAs achieved from the post-panning populations is being lost during purification steps. Therefore elimination of the propagation step would require a more efficient high-throughput method of isolation and purification of large phage DNA molecules. Potential problems with this step are treated in section 3.5.9 below.

An additional experiment was performed in order to evaluate whether the microarray approach is sensitive enough to detect protein enrichment, when a higher dilution of model phages was applied. For this purpose, phages displaying the three model domains were supplemented to the T7Select10-3b library, each at a ratio of only 1 in 10^6 pfu. The phage mixture was used for one round of affinity selection over the pep-E ligand. The eluted phage population was propagated and 200 µl of the post- and pre-panning mixtures were taken for a phage DNA template preparation as described in 2.2.3.5. The aliquots of 65 µl of both templates were taken for PCR labeling reaction. The pre- and post-panning DNA inserts were labeled with the Cy5 and Cy3 dyes, respectively. Additionally, PCR labeling of a control DNA insert coding for the Fas antigen was carried out in two separate reactions for Cy3 and Cy5. A simple DNA microarray biochip was prepared using a contact printing microarrayer (see 2.2.5.2). DNA inserts coding for the three model domains, control Fas and two other irrelevant DNA inserts (Interleukin-4 and Ezrin), each concentrated to 140-160 ng/µl, were printed onto super amine slides (Telechem). The hybridization mixture was prepared by mixing equal amounts of labeled pre- and post-panning DNA (2 µg, each) and 500 ng of each of control Fas inserts. The mixture was hybridized to the microarray under a cover-slip glass as described in 2.2.5.8. The results achieved for phages isolated from the pep-E, as well as a schematic representation of the whole process are shown in (Fig 3.5.5). Although a strong dilution of all model phages within the pre-panning population, those, displaying the EVH1 domain were significantly enriched on the corresponding peptide, which was detected as the green signal emitted from the EVH1 DNA on the microarray. The DNA probes coding for the WW/F65 and WW/YAP domains didn't show a green signal here, which confirms a lack of interaction of both domains with this peptide ligand. The lack of a red signal from these probes as well as from the EVH1 probe is due to the high dilution of these model phages in the pre-panning mixture and thus too little of their Cy5 labeled DNA inserts. This is much better visible when the array image was analyzed with inactivation of the channel for the green fluorescence (Fig 3.5.6). Nearly equal amounts of Cy5- and Cy3-labeled control

Fas-DNA hybridized to its corresponding DNA on the array, resulting in a yellow color (Fig 3.5.5). DNA probes corresponding to IL-4 and Ezr showed no signals because such phages were not supplemented to the pre-panning mixture and were obviously not present or not enough abundant in the T7 phage library.

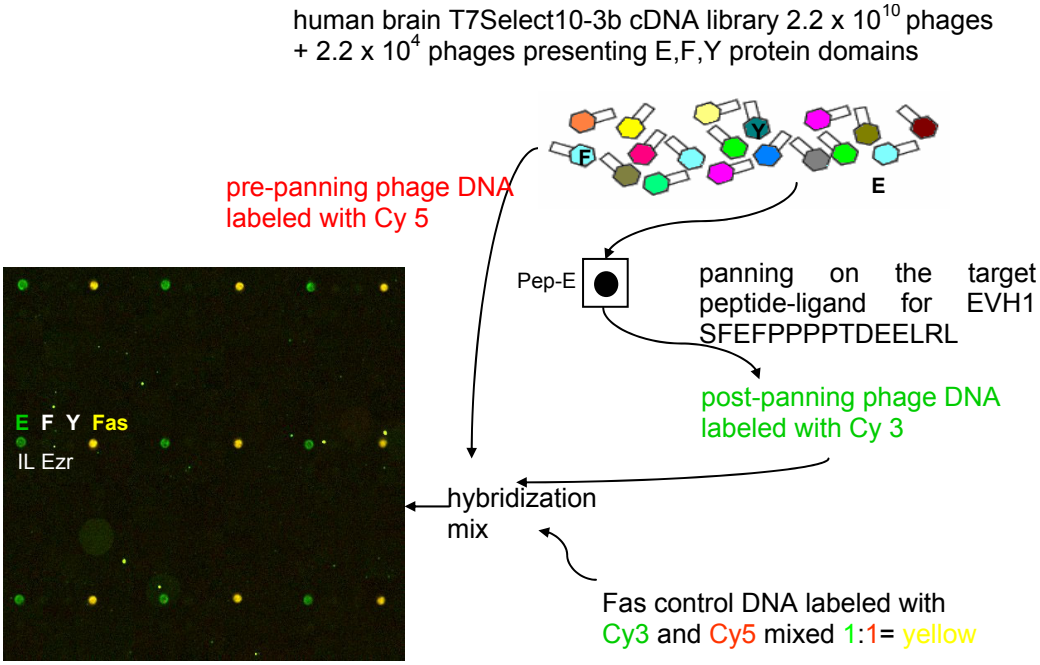


Fig 3.5.5: Schematic picture showing the panning process for the phage library supplemented with the model domain phages, on the peptide ligand for the EVH1 domain (pep-E) and the resulting image obtained from the differential hybridization of the pre- and post-panning phage library DNA to the DNA probes printed onto a glass slide. Short name code: EVH1 (E), WW/FE65 (F); WW/YAP (Y); Ezrin (Ezr); Fas antigen (Fas); Interleukin-4 (IL).

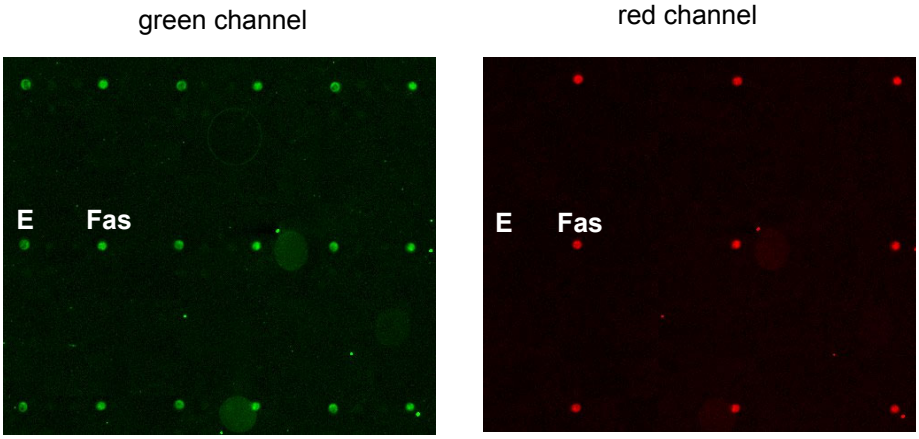


Fig 3.5.6: The image of Fig 3.5-5, generated using the green and red fluorescent channel separately.

3.5.3 Choice of the full length cDNA probe library

For a successful application of microarrays to the screening format described above, the collection of cDNA sequences to be printed on a biochip is ideally a replicate of the population of cDNA inserts cloned into the T7 phage chromosome. Several libraries of clones containing human cDNA sequences are provided by the RZPD in Berlin. The E. *coli* clone library containing full length cDNA sequences originating from the human brain region of amygdala, which were cloned into the pSPORT1 phagemid vector, was considered as the most suitable one for this approach. Both, the amygdala library and the T7 phage display library were generated from mRNAs isolated from human brain. Therefore a high percent of the same genes present in both libraries was expected. Additionally, a length of cDNA inserts from the amygdala region (up to 3.6 kb), was expected to cover even the largest inserts from the T7 phage library (up to 1.5 kb). Moreover, a large number of cDNA sequences present within this library is already sequenced, allowing an easy identification of sequences of enriched phages, without sequencing of clones corresponding to a DNA printed on a biochip. A part of this library, 6144 bacterial clones in 16 plates (384 wells), was purchased from the RZPD. A replica of each micro-plate was prepared immediately after receipt to reduce the risk of contamination of the original library. Clones were replicated using 96 pin tool into four 96 deep well blocks filled with LB and grown as described in 2.2.1. Aliquots of 30 µl were transferred into 96 well cryo-resistant plates, supplemented with glycerol (20 % f.c) and frizzed at -80°C. Pins of the replication tool fit into every second well of the 384 micro-plate, therefore each from the four 96 replica plates was labeled with respect to the corresponding position for the first pin in the left upper corner of the 384 source plate (i.e. A1, A2, B1, B2; Z-type arrangement). Additionally each replica contained a number of the original RZPD plate (i.e. PL1A1= plate 1 position A1).

3.5.4 Comparison of plasmid and PCR probes spotted on a glass surface

Fabrication of a standard cDNA microarray according to standard literature protocols would require a high-throughput PCR amplification of a library of DNA inserts and printing of purified PCR products on a glass slide. High-throughput PCR amplification is however the most laborious, difficult, time and cost consuming step in the whole procedure of cDNA array preparation (Botwell *et al.*, 2002). Thus, a much less complicated process, utilizing plasmid preparations for printing, would be very much preferable. The main question to be answered was, whether the plasmid array is sensitive enough to detect phages enriched on target peptides? Therefore studies comparing the performance of arrayed plasmid and PCR probes were carried out. Suitable constructs containing two plasmid vectors and cloned inserts encoding the three model domains (EVH1, WW/YAP, WW/FE65), as well as the other three

control proteins (Fas, IL-4 and Ezrin), were engineered. Two different plasmids were chosen, the pOBT7 and pCMV-sport6, having significantly different sizes, 1815 bp and 4396 bp, respectively. Since a lower number of larger DNA molecules can be deposited on a single spot compared to smaller DNA molecules, printing of differently sized plasmids was expected to elucidate possible problems of probe fluorescence intensities that might be caused by different amounts of spot localized DNA molecules. The DNA sequences of inserts were transferred to the plasmid vectors from the T7 phage constructs prepared previously (see Bialek *et al.,* 2003). Inserts were PCR amplified from the T7Select103b phage vector using a suitable set of primers (see Table 2.1.6.1). Plasmids were purified from bacterial cultures using Qiagen's plasmid purification columns. Both were first cut with the *Eco*RI and *Xho*I restriction enzymes, dephosphorylated (only plasmids), purified and ligated (see methods). Subsequently, the *E. coli* DH10B bacteria were transfected with the appropriate constructs and propagated. Plasmids were purified as above. All purified plasmid constructs were concentrated to 300 ng/µl and transferred into a 96 well spotting plate. Additionally, cloned DNA inserts were PCR amplified from both plasmids, digested with *Eco*RI and *Hind*III enzymes, L.M.P purified and concentrated using vacuum evaporator. PCR probes obtained from pOBT7 plasmid were recovered with the spotting buffer (3 x SSC, 1.5 M betaine) giving a yield ranging from 250 to 300 ng/µl. Inserts amplified from the pCMV-sport6 plasmid were recovered in a very low amounts, ranging from 40 to 100 ng/µl for all inserts except Fas (190 ng/µl). The reason for that was most probably over-drying of DNA. Because of their low concentrations, these probes were not taken into account during final analysis of arrayed probes. Both, plasmid constructs and PCR products were printed on the amine functionalized glass slides. Phage panning was carried out over the pep-E spot, using mixture of T7 phage library (titer 7.3 x 10^{10} pfu/ml), supplemented with phages displaying the three model domains, each containing 7.3 x 10^5 pfu (dilution ratio 1:10^5). As described before, pre- and post-panning populations were propagated, PCR amplified and labeled with Cy3 and Cy5, respectively. DNAs from both mixtures were digested with the *Eco*RI and *Hind*III restriction enzymes, and purified. Following concentration measurement, both DNA populations were mixed together (1.32 µg each). Unlike as described in the above experiments, five genes instead of three were used as controls in this case. These include DNAs of WW/FE65, WW/YAP, Fas, IL-4 and Ezrin (see below). Two separate PCR reactions, each containing mixtures of the five phage templates with the above inserts were carried out simultaneously to label them with the Cy3 and Cy5 dyes. The PCR products were digested and purified as above. Equal amounts of differentially labeled inserts were mixed together in equal amount. An aliquot containing 2 µg of such prepared mixture was added to the mixture, which was applied on the PCR-plasmid microarray for hybridization. The results of hybridization to the printed PCR probes could be foreseen on the basis of the previous

experiments. However, nothing is known about the efficiency of hybridization to plasmid DNA probes. Application of as many as five controls thought to help investigating hybridization to plasmid probes in the case these were not sensitive enough to detect phage enrichment. Theoretically, DNA of externally supplemented controls should be more abundant within the hybridization mixture (200 ng each in this case) compared to DNAs of enriched phages. Therefore at least these over-represented control genes should efficiently hybridize to the plasmid probes. Moreover, sizes of the three standard controls (Fas, IL-4, Ezrin) range from 500 to 1100 bp, whereas the two additional inserts used here as controls (WW/FE65 and WW/YAP) are at least two times smaller, 180 bp and 240 bp respectively. This could help to observe a possible relationship between sizes of the inserts and hybridization efficiencies to plasmid probes. The results obtained, showed that despite their high concentration, neither spots containing the smaller pOBT7 or the bigger pCMVsport-6 plasmids gave sufficient fluorescent signals (Fig 3.5.7).

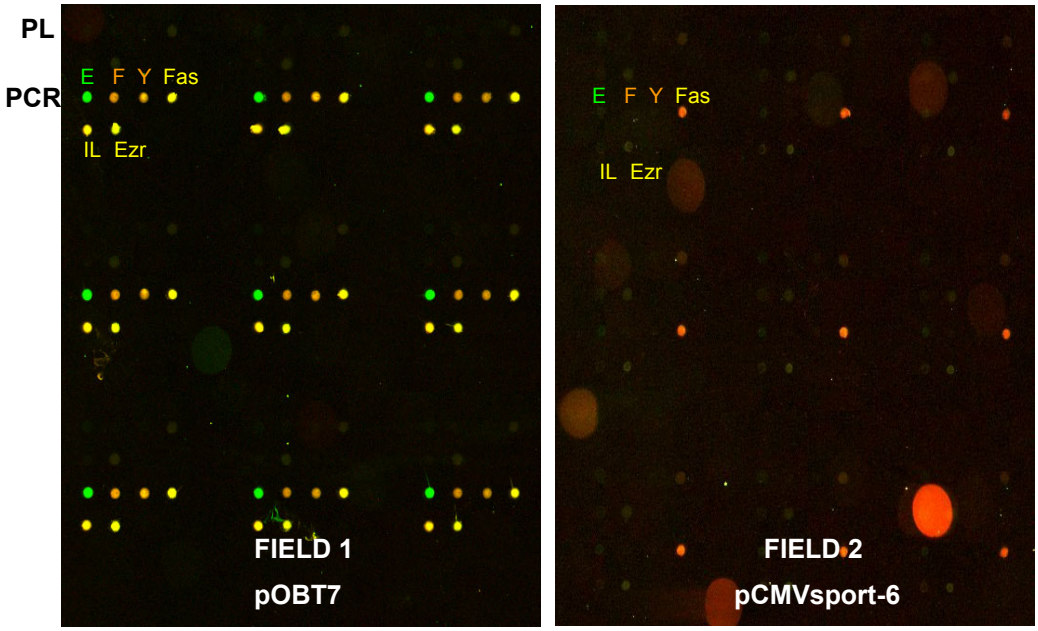


Fig 3.5.7: Comparison of phage enrichment detection sensitivities using PCR and plasmid arrays. Differential hybridization of the pre- and post-panning phage library DNA to the EVH1 domain-specific DNA probe printed onto a glass slide. Field 1: contains the pOBT7 plasmid probes and PCR probes amplified from pOBT7 constructs. Field 2: contains the pCMV-sport6 plasmid probes and PCR probes amplified from the pCMV-sport6 constructs. Short name code: plasmids (PL), PCR products (PCR), EVH1 (E), WW/FE65 (F); WW/YAP (Y); Ezrin (Ezr); Fas antigen (Fas); Interleukin 4 (IL). Each set of probes is printed in three replicates.

As expected, all PCR probes amplified from the pCMV-sport6 constructs that were printed on the field 2 (see Fig 3.5.7) didn't show any significant signals as well, most probably because of technical problems as mentioned above. In contrast to these observations, signals achieved from the PCR probes amplified from the pOBT7 constructs were well visible and satisfactory (see field 1). The affinity enrichment of phages displaying the EVH1 domain on its peptide ligand was very well visible as a green signal obtained from the spotted insert coding for the EVH1. Spots containing the WW/FE65 and WW/YAP products emitted slightly orange signals, showing that more inserts labeled with the Cy5 hybridized to these probes. This result can be explained by the fact that beside both WW/FE65 and WW/YAP control inserts labeled with both dyes, there were additional inserts present within the hybridization mixture, which origin from the pre-panning DNA population. Therefore the Cy5 labeled DNAs were more abundant within the entire population, hybridizing to the probes and contributing to the fluorescent signal. The other control PCR probes showed yellow color, demonstrating an equal hybridization of control inserts labeled with both dyes. The absence of signals form the plasmid probes was quite disappointing. Since the cloned inserts represent only a small part of the entire plasmid DNA (insert: plasmid ratios from 1:3 to 1:20), the number of molecules which are available for hybridization is correspondingly lower than the number of PCR probe molecules spotted on the glass slide in the same or similar optically measured DNA-concentration. These differences obviously results in a relatively weak signals from the plasmid probes. Moreover, due to a super-helical structure of plasmids and their larger size, many inserts could theoretically be hidden (i.e. buried) and therefore not available for hybridization with the labeled inserts. One of the possibilities to avoid such situation and achieve better accessibility of the inserts would be a linearization of plasmids to be spotted by application of restriction enzymes. This additional and laborious step however, would significantly increase costs of the entire process and therefore was not considered here. The comparison experiments performed here clearly showed that only PCR probes can be used. Therefore a fabrication of a PCR microarray will be a subject of the following sections.

3.5.5 Large scale preparation and spotting of PCR products

The next goal then was the establishment of an efficient PCR amplification of inserts varying considerably in size (0.5 kbp – 3.6 kbp) applicable top most of the clones of the amygdala clone library. The PCR was optimized using a small set of randomly chosen clones, before the whole clone library was amplified (data not shown). For this purpose several PCR ingredients and their concentrations as well as program parameters were stepwise varied. Concentrations of different primers varied from 10 to 20 pmol, magnesium ions ($MgCl_2$ or $MgSO_4$) from 1 mM to 4.5 mM, nucleotides from 0.2 to 0.4 mM. Program parameters varied

as follows: initial denaturation step (94°C) 1-5 min, primer annealing temperatures (45-65°C), annealing times (20-50 s), elongation (1.30-5 min) and number of cycles (20-40). Experimentally established primer pair and optimal PCR conditions are shown in 2.1.6.1 and 2.2.3.11, respectively. Since the high-throughput PCR requires relatively large amounts of expensive enzymes, the effort was taken to produce *Taq* DNA polymerase in-house using an *E. coli* expression system (see 2.2.7). Its quality was proven by PCR amplification of insert coding for the EVH1-domain using various dilutions of the enzyme and the primer set (M13 for 23, M13 rev 30) shown in Table 2.1.6.1. The previously described pCMVsport-6+EVH1 construct was used as a template for amplification. Similar amplification was carried out using a commercially available *Taq* DNA polymerase, purchased from NEB. It was revealed that the performance of self-made polymerase was even better than the performance of the commercial enzyme (Fig 3.5.8).

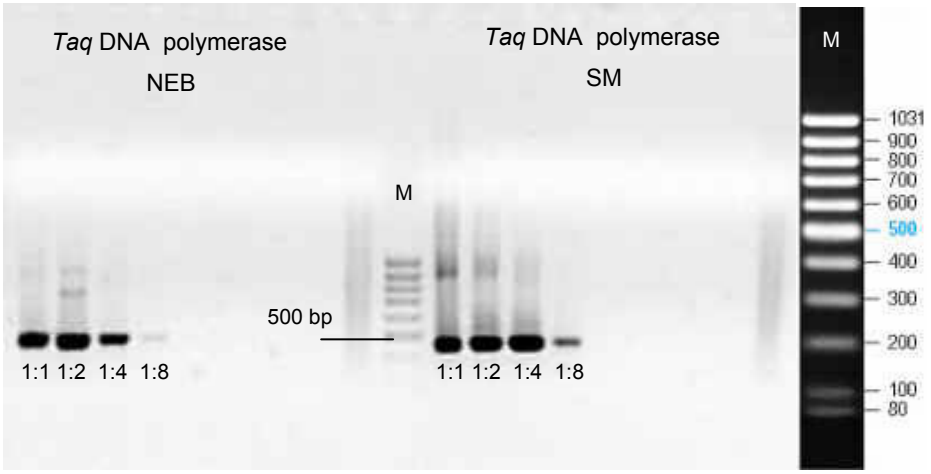


Fig 3.5.8: Comparison of the performances of commercial NEB and self made (SM) *Taq* DNA polymerases. Ethidium bromide stained 1 % AGE analysis of EVH1 DNA inserts. C) Control PCR without template supplement, 1:1-1:8) Enzyme dilutions, M) MassRuler™ DNA Ladder, Low Range, ready-to-use (MBI Fermentas).

The most important aspect during preparation of DNA probes to be spotted is the quality and yield of amplified inserts. *Taq* DNA polymerase has been reported to make frequent mismatch mistakes when incorporating new bases into a strand (Huang *et al*., 1992. Ayyadevara *et al.,* 2000). An efficient amplification of long DNA fragments can be improved via application of high fidelity polymerases or ready to use kits containing mixtures of enzymes with different performance provided by several companies (Quiagen, Stratagene, Fermentas etc.). The addition of a low amount of a proofreading enzyme (e.g. *Pfu* DNA polymerase) to PCR reaction mixtures has been proposed to improve the performance of non-proofreading polymerases (e.g. *Taq* DNA polymerase) by correcting mismatches

introduced during PCR, which otherwise prevent the efficient synthesis of full-length products (Barnes, 1994, Cline *et al.,* 1996). Additionally, PCR enhancing additives (DMSO, glycerol, BSA, betaine) have been reported to improve amplification of some DNA inserts (Chen and Janes, 2002, Frackman *et al.,* 1998). Test PCRs of four randomly chosen amygdala library clones of insert sizes ranging from ~1.5 to ~2.5 kb, were carried out in order to check whether an enzyme mixture containing self-made *Taq* DNA polymerase and *Pfu* DNA polymerase (MBI Fermentas), as well as the addition of betaine would increase a yield and quality of amplified inserts. Both enzymes were supplemented to a PCR reaction mixture at ratio 10 U *Taq*: 1U *Pfu.* Three similar reactions were performed as described in 2.2.3.11 in which different elongation times were applied (2.0 - 3 min). In comparison to the *Taq* polymerase alone, the combination of both enzymes resulted in higher product yields, observed for all three cases (Fig 3.5.9).

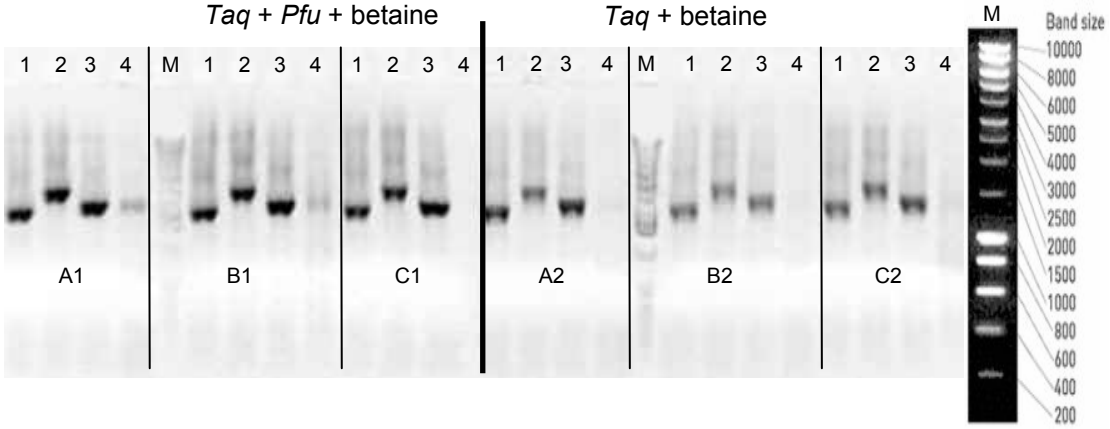


Fig 3.5.9: Performance comparison of the SM *Taq* DNA polymerase alone and in combination with a proofreading *Pfu* polymerase. Ethidium bromide stained 1 % AGE analysis. 1-4) clone No, A1, A2) PCR with elongation time 2.3 min, B1, B2) PCR with elongation time 2 min, C1, C2) PCR with elongation time 3 min, M) DNA marker SMART (Eurogentec).

Subsequently, PCRs of the same inserts (2.2.3.11), but using four different concentrations of magnesium sulfate were carried out to check whether such variation would improve amplification of the most problematic inserts. However, the significant increase in its yield was not observed, suggesting that some inserts are difficult to be amplified, even when sophisticated conditions are applied (see Fig 3.5.10). Additionally, the influence of betaine enhancer was checked by amplification of the other four randomly chosen inserts. The reactions were carried out similarly as for the above fragments with and without addition of betaine and using different salt concentrations (1 mM and 4 mM). Positive effects were noticed for all amplified fragments (Fig 3.5.11). Again, 4 mM concentration of magnesium

sulphate resulted in the highest DNA yields. On the other hand the insert No.2 seemed to amplify better at lower salt concentration (1 mM). This suggested that uniform high-throughput amplification of all library inserts is very difficult and most probably a certain percent of inserts from the whole library will not be efficiently amplified.

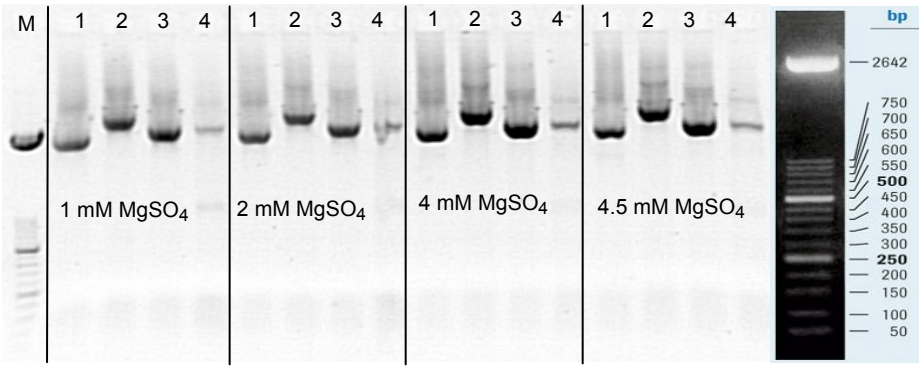


Fig 3.5.10: Comparison of insert amplification using different concentrations of $MgSO_4$. Ethidium bromide stained 1 % AGE analysis. All probes were amplified using *Taq*/*Pfu* mix and 1 M betaine. 1-4) amplified inserts. M) DNA marker MXII (Boehringer Mannheim).

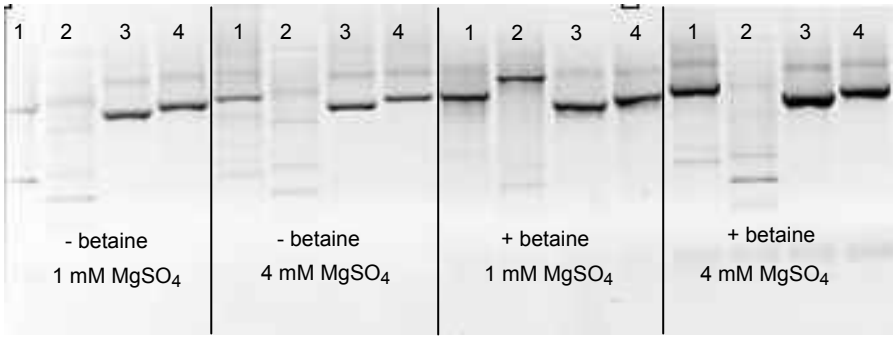


Fig 3.5.11: Comparison of insert amplifications with and without addition of betaine. Ethidium bromide stained 1 % AGE analysis. All probes were amplified using *Taq/Pfu* mix. 1-4) amplified inserts.

The above PCR optimization experiments were carried out using purified plasmids as a template. Although, this is certainly preferred for PCR application in general (Holloway *et al.,* 2002), purification of thousands of plasmids from the amygdala library would be challenging and costly. Alternative sources of DNA templates are: (i) lysates prepared from fresh overnight bacterial cultures, (ii) directly used bacterial culture (1 µl) or (iii) single bacterial colony. The lysates are prepared by heating (95°C) mixtures containing 5 µl bacterial in 95 µl water, followed by centrifugation. Usage of single colony or bacterial culture requires increased time of initial denaturation step during PCR to lyse bacteria *in situ* (Botwell *et al.,* 2002). To check whether the alternative templates will be sufficient for PCR amplification,

eight randomly chosen amygdala clones were amplified, from centrifuged bacterial lysates (i), bacterial cultures (ii) or bacterial colonies (iii) and compared to purified plasmid templates. PCRs for plasmids (~50 ng) and bacterial lysates (i), (vol 4 µl), were performed exactly as described in section 2.2.3.11. The initial denaturation step was extended up to 5 min for colony PCR (iii) and bacterial cultures (ii). Instead of using expensive commercial kits, plasmids used in this experiment were purified using the shortened alkaline lysis protocol described in 2.2.3.6. Only two products were obtained in the case of lysates (i) and colonies (i), similarly, amplification from bacterial cultures (ii) was very weak (not shown). However, all inserts amplified from purified plasmids gave satisfactory yields (Fig 3.5.12). This experiment proved also that plasmids preparations achieved by shortened alkaline lysis protocol, are pure enough to be used for efficient PCR amplification.

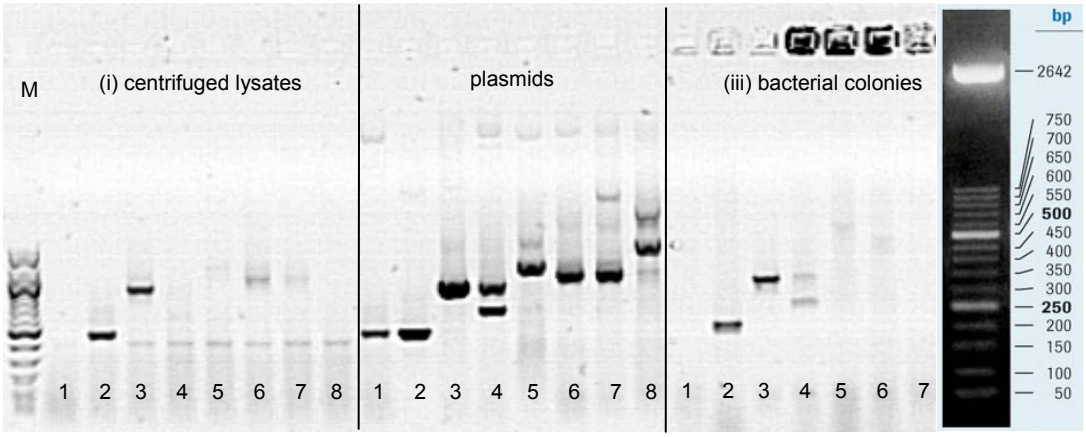


Fig 3.5.12: Comparison of insert amplifications when, centrifuged lysates (i), plasmids and bacterial colonies (iii) were used as a source of DNA template. Ethidium bromide stained 1 % AGE analysis. 1-8) amplified clone inserts, M) DNA marker MXII (Boehringer Mannheim)

Additional PCR tests were performed utilizing centrifuged bacterial lysates, where different volumes of fresh bacterial cultures were taken for their preparation (5 µl + 95 µl H_2O, 200 µl, 400 µl). Four micro liters aliquot from each lysate was taken for PCR. No products were obtained after amplification from the lysate containing 5 µl of bacteria. Five among eight inserts were amplified from lysates containing 200 and 400 µl of bacteria (Fig 3.5.13). However, none of the remaining lysates gave as satisfactory results as were obtained from purified plasmids above. Finally, plasmid preparations by the shortened alkaline lysis method were chosen as a source of templates for future high-throughput amplification of inserts.

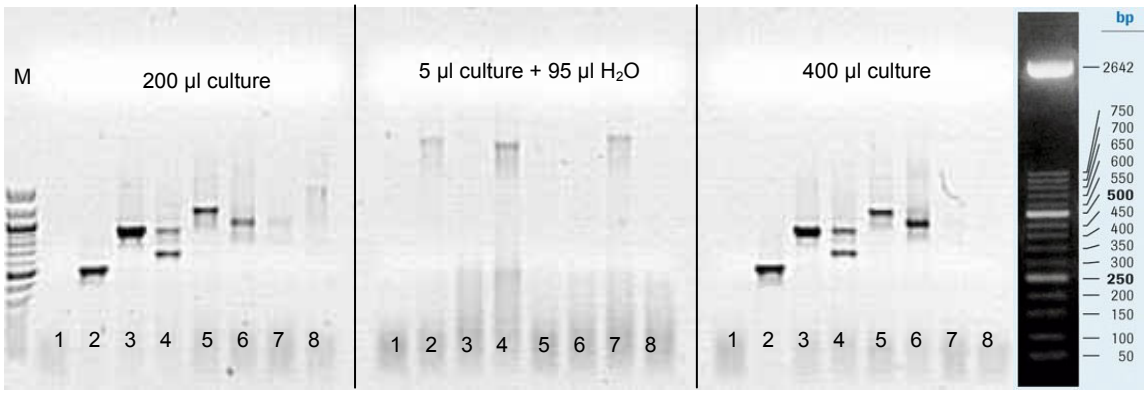
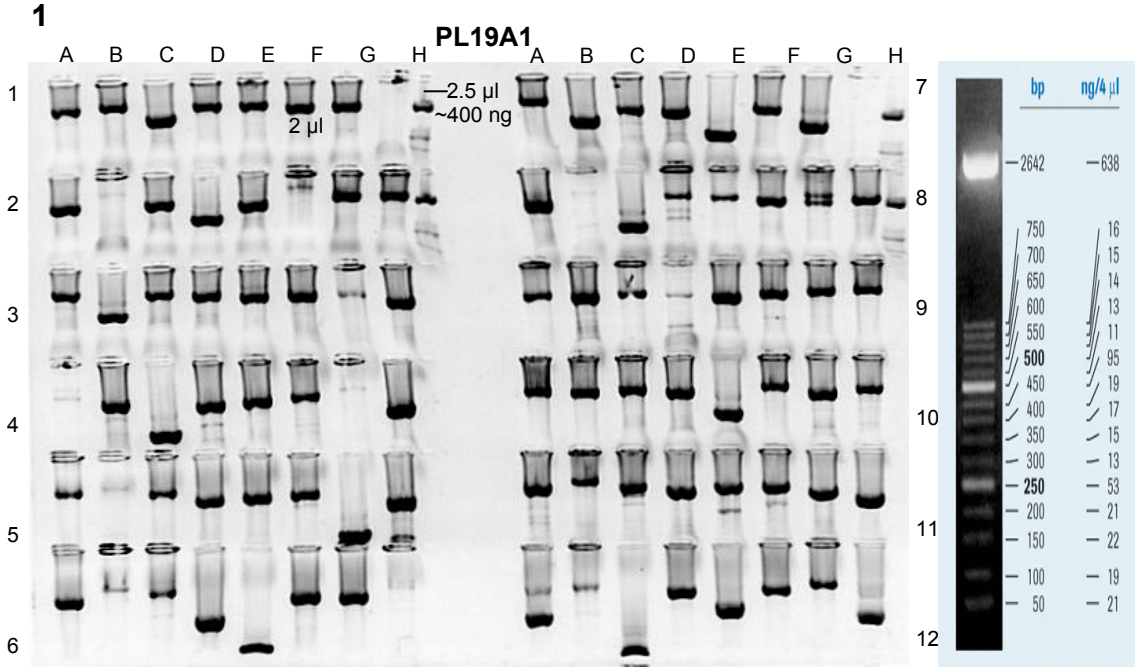


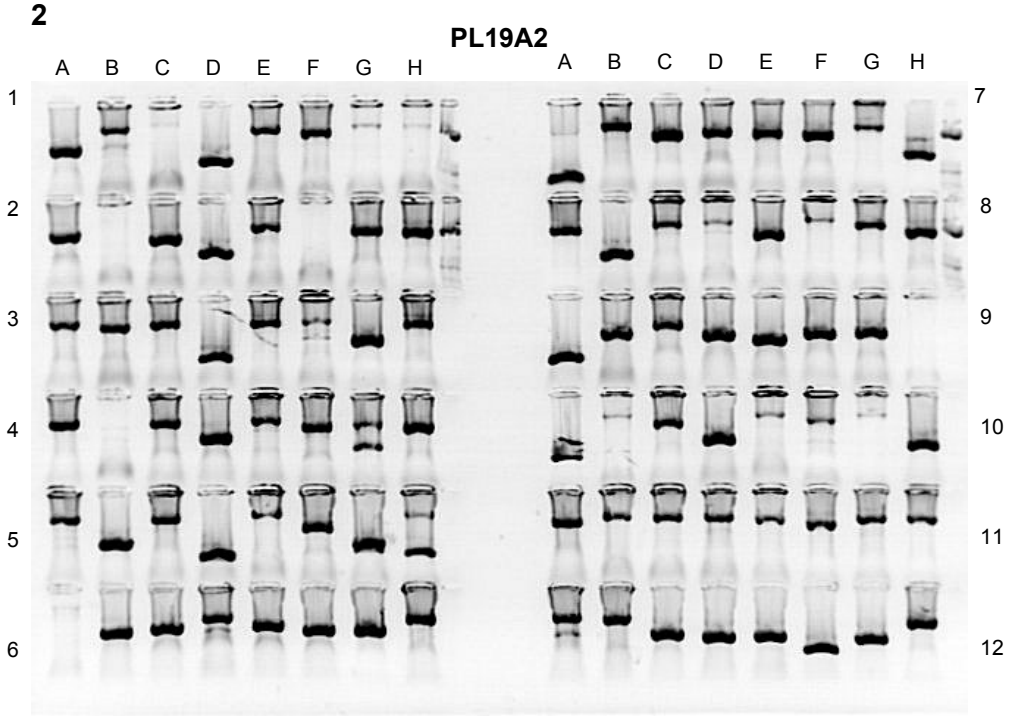
Fig 3.5.13: Amplifications of inserts using templates obtained from centrifuged lysates containing different amounts of bacteria. Ethidium bromide stained 1% AGE analysis. 1-8) amplified clone inserts, M) DNA marker MXIII (Boehringer Mannheim).

From the amygdala library clones, bacterial cultures were prepared by transferring the inoculums from previously prepared 96 well replica plates into 96 deep well culturing blocks, using 96 pin replication tools and growing bacteria overnight (see 2.2.1). The modified alkaline lysis protocol was applied for plasmid purification as described in 2.2.3.6. PCRs were performed in 96 well plates using MJ Research Thermal Cycler instruments as described in 2.2.3.11. Utilization of 96 well plates allowed for a more efficient performance of the process by using a multi-channel pipette. Steps for the separation of the DNA fragments from enzymes, nucleotides, primers and other PCR components and the adjustment of an appropriate concentration for ready to spot fragments were optimized. Amplification of insert from each plate was carried out four times giving a total amount of 24,576 reactions. In order to reduce PCR purification costs, DNA precipitation with isopropanol was performed as described in 2.2.3.1 (Diehl *et al.,* 2002; Sambrook and Russel 2001). Purified inserts were collected, precipitated once again, air dried and recovered via addition of 15 µl of spotting buffer (3 x SSC, 1.5 M betaine). All PCR products were controlled electrophoretically using 96 pocket agarose gels. Electrophoresis was carried out with a high-throughput Electro-Fast WIDE system (AB Gene). Example images of amplified inserts from library plate No.19 (384 wells) are shown in Fig 3.5.14. Concentrations of PCR products were estimated approximately via comparison of band intensities with intensities of DNA molecular mass marker and ranged from 160-250 ng/µl. An average percentage of PCR failure was estimated as 15 %, from the number of failures among 96 reactions.

1


PL19A1
A B C D E F G H
1
2
3
4
5
6
2 µl
—2.5 µl
~400 ng
7
8
9
10
11
12
bp
ng/4 µl
2642
638
750
700
650
600
550
500
450
400
350
300
250
200
150
100
50
16
15
14
13
11
95
19
17
15
13
53
21
22
19
21

2


PL19A2
A B C D E F G H
1
2
3
4
5
6
7
8
9
10
11
12

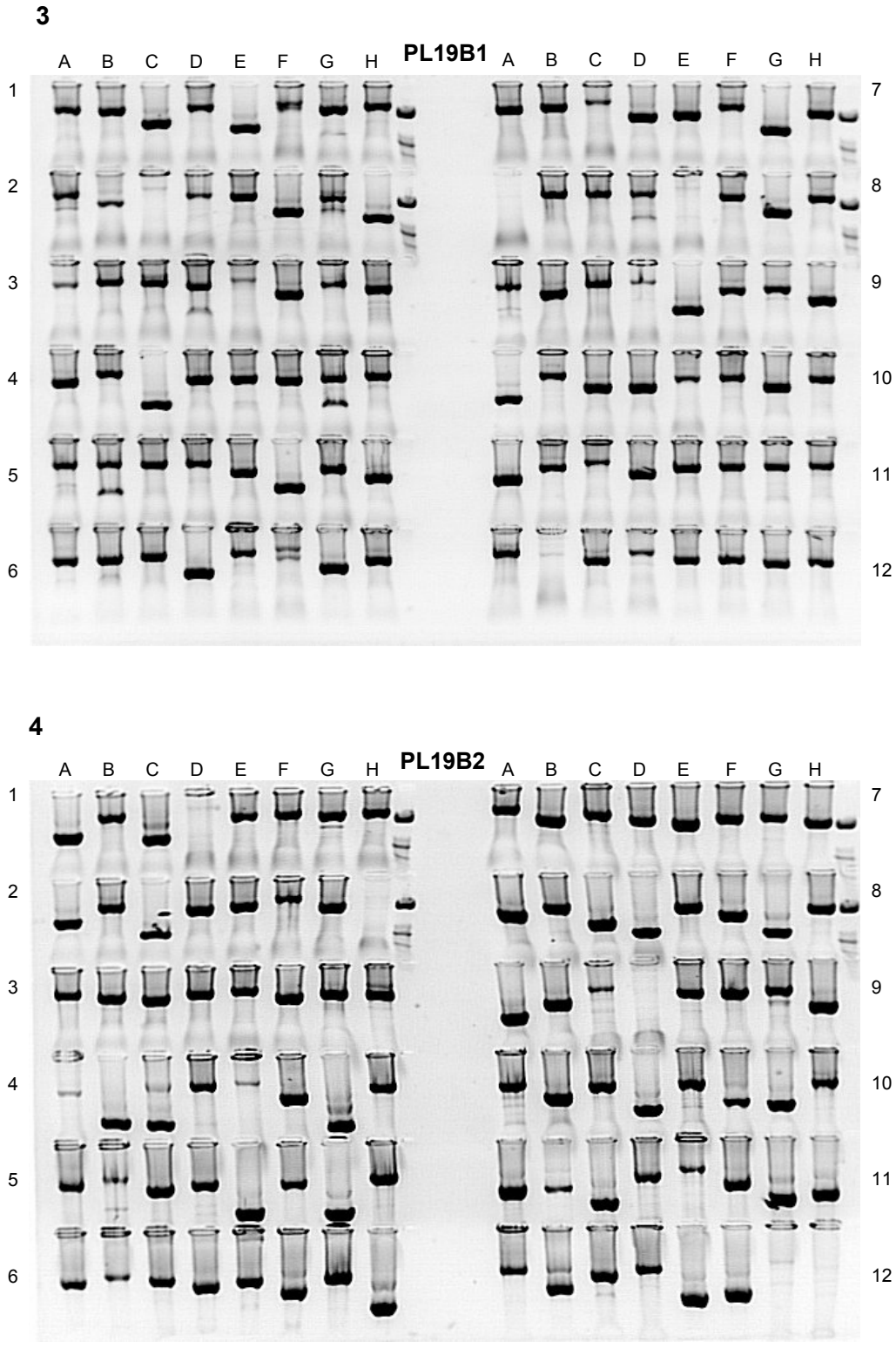


Fig 3.5.14: Ethidium bromide stained 1 % AGE analysis of PCR amplified inserts from the amygdala library plate No.19 (384 wells = 4 x 96). A-H and 1-12) positions in 96 well PCR plate. M) Molecular mass DNA marker MXII (Boehringer Mannheim).

Purified PCR fragments were transferred into 384 wells spotting microplates and printed onto the super amine and amine silanized slides as described in (2.2.5.2, appendix 1). Quality of the printed DNA microarrays was checked via hybridization of a selected slide with 15mer random primers labeled with Cy3 dye (Fig 3.5.15).

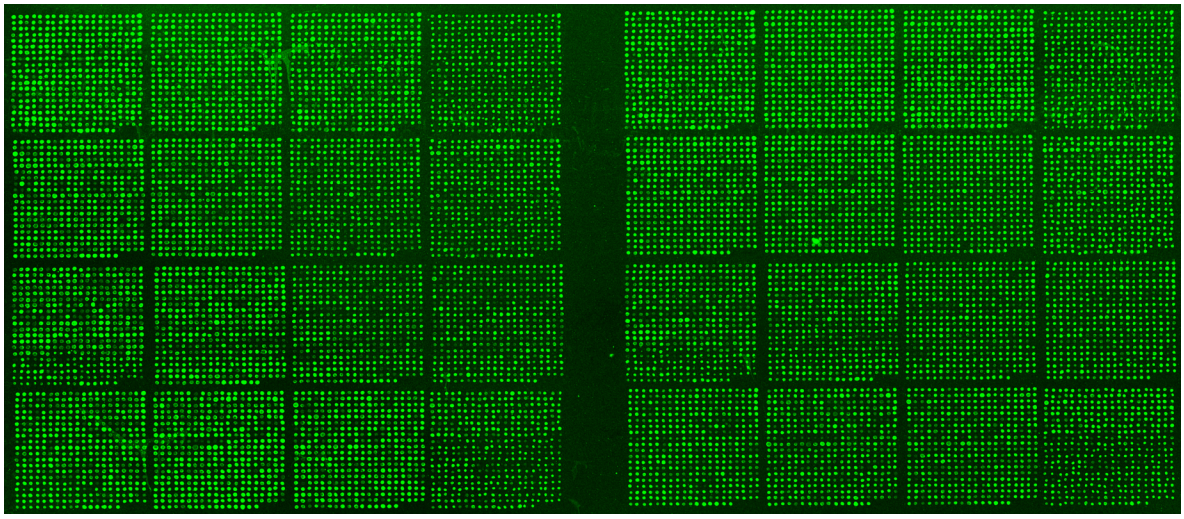


Fig 3.5.15: Quality check of amygdala microarray by hybridization with Cy3 labeled random primers. Two identical replicas were spotted on the glass slide.

3.5.6 Functional analysis of microarray experiments

To check detection performance of the amygdala array, panning experiment utilizing phages displaying the three model protein domains over the pep-E was carried out. Model phages were supplemented to human brain T7 phage display library (5.4×10^{10} pfu) at ratio $1:10^6$ to give a final concentration of 5.4×10^4 pfu, each. Following phage elution, propagation and DNA labeling, equal amounts of pre and post-panning DNA and Fas control DNA were mixed together and hybridized to the amygdala biochip using the automatic Lucidea hybridization station (Pharmacia). Despite low signal intensities in general, as well as unsatisfactory quality of the array image due to a high Cy3 green fluorescence background, a very strong green signals from the EVH1 DNA probe indicating its enrichment and a yellow from the Fas control was well visible (Fig 3.5.16). As expected, the remaining model protein domains didn't bind to the pep-E, therefore no other significant signals indicating phage enrichment were observed. The overall reddish signals visible on the slide were a result of hybridization of Cy5 labeled pre-panning phage DNAs to their complementary sequences printed on the glass surface. This demonstrated that the representation of genes on the amygdala array correspond well with a major part of genes within human brain T7 phage library. Except for

the signals emitted by the EVH1 probes, no other green signals indicating enrichment of any other proteins on the pep-E ligand, were observed among the remaining 6144 amygdala probes on the array.

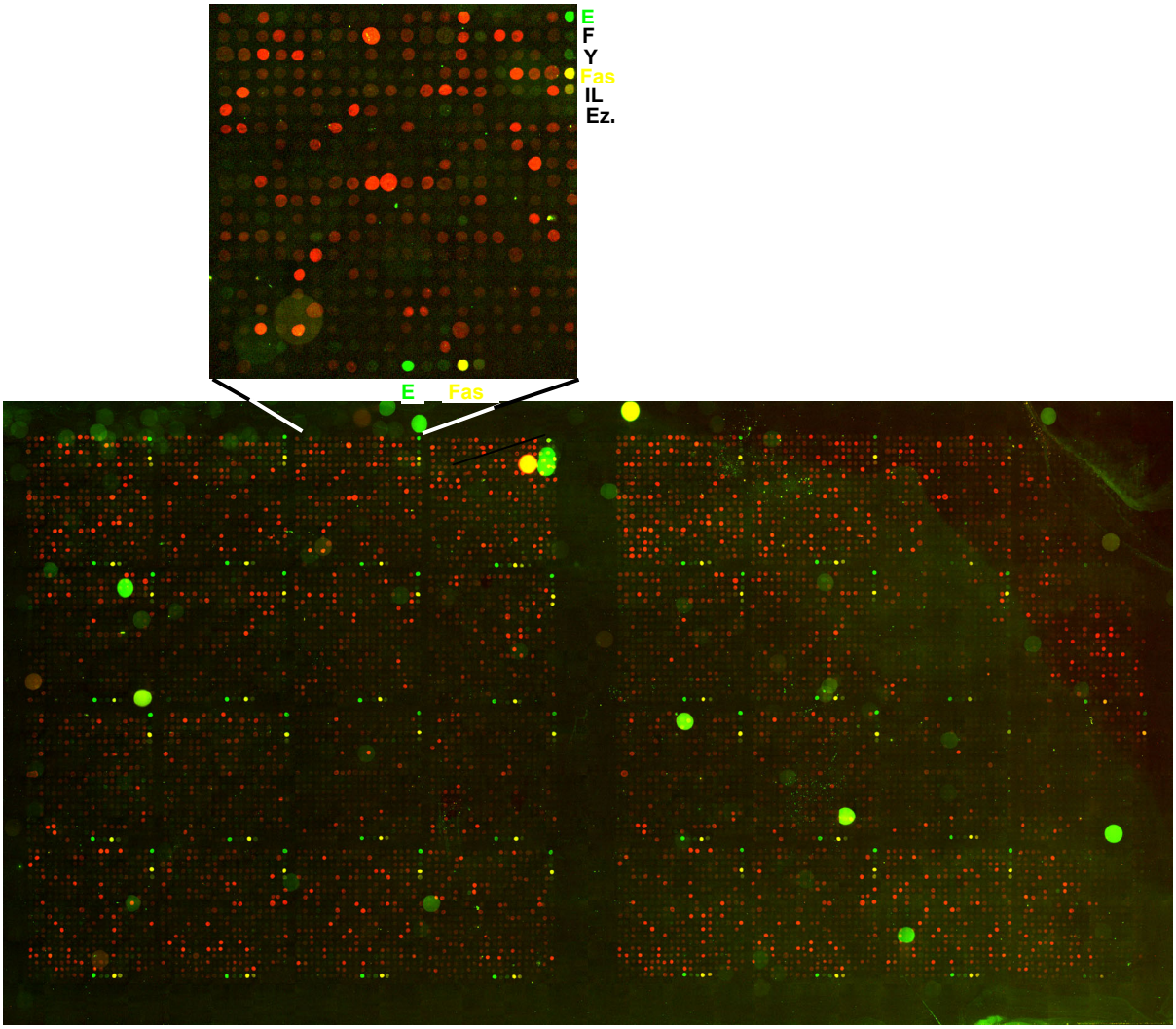


Fig 3.5.16: Results obtained from the differential hybridization of the pre- and post-panning DNA populations to the EVH DNA probe present on the amygdala microarray. Short name code: EVH1 (E), WW/FE65 (F); WW/YAP (Y); Ezrin (Ezr); Fas antigen (Fas); Interleukin-4 (IL).

Unfortunately, the data form many other experiments could not be evaluated because of a high green fluorescence background, which caused difficulties in measuring the true spot signals and therefore distorted the microarray analysis. This problem was especially observed for slides that were chemically blocked and washed with dichloroethane (DCE). It suggested that residuals of the blocking substances on the glass surface are attractive for the Cy3 dye. Additionally, many dust particles decreased data quality. Some example array sections with high background Cy3 fluorescence are shown in Fig 3.5.17.

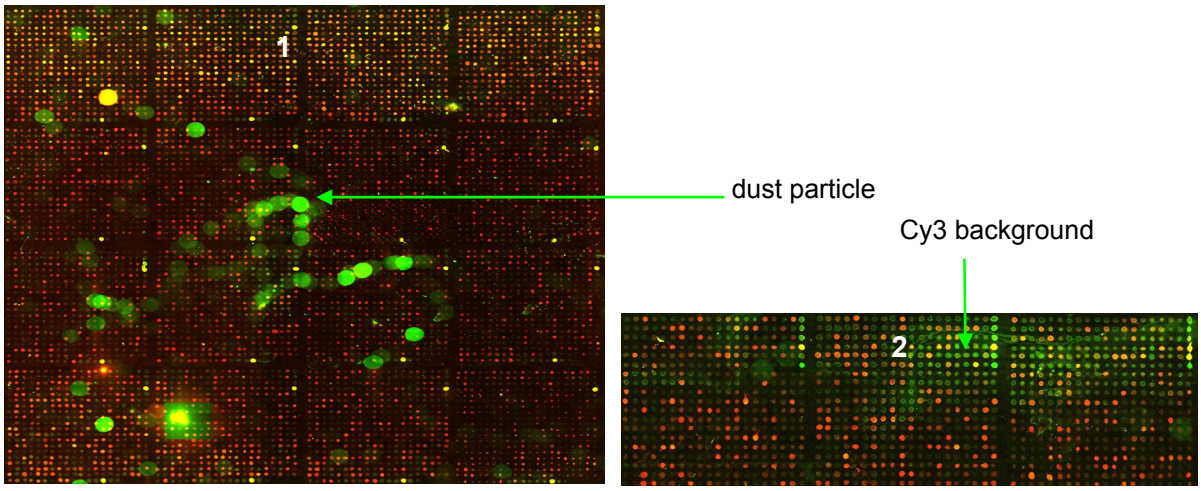


Fig 3.5.17: Images 1 and 2 showing problematic artifacts with the Cy3-dye fluorescence.

Analysis of the same arrays using separate laser channels showed that Cy5 red background fluorescence was much lower in comparison to Cy3 fluorescence. Interestingly, a majority of the dust particles present on the array No1 showed fluorescence in the wavelength range of the Cy3-dye only. Similarly, contaminating background fluorescence on the slide No2 was visible only when the green laser channel was used (Fig 3.5.18). This suggested that the application of Cy5-dye only would allow for more correct measurements of even weak hybridization signals. Additionally, one dye system would eliminate problems related to biases in dye incorporation during PCR labeling procedure (Yu *et al.,* 2002, Richter *et al.,* 2002), as well as their different environmental stabilities (Fare *et al*. 2003).

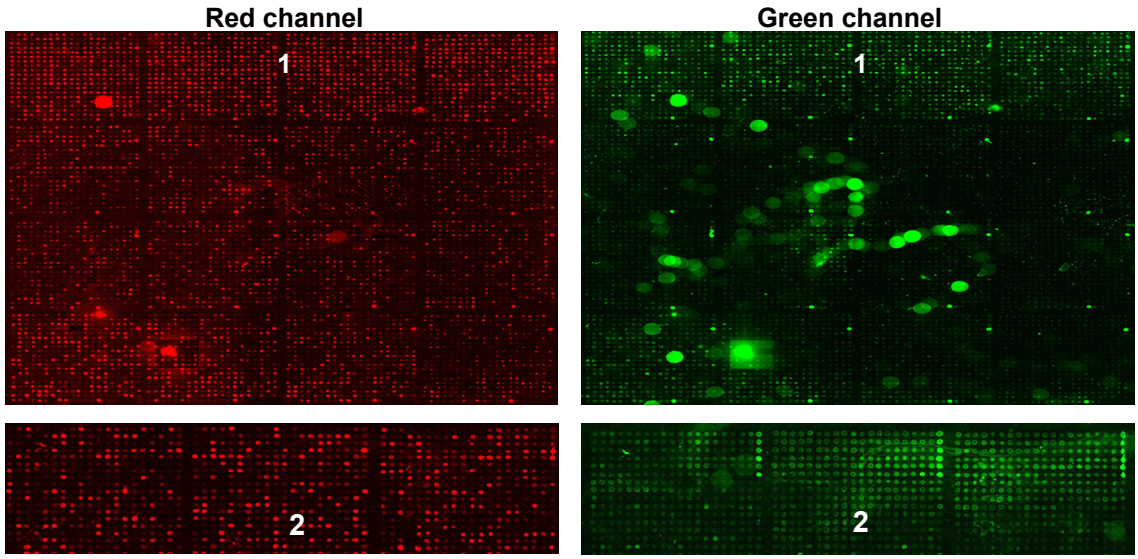


Fig 3.5.18: Images of Fig 3.5.17 generated using the green and red channel separately.

Introduction of the one dye detection system required minor modifications of the existing protocol. Two microarray biochips had to be used, instead of one only, for separate hybridizations with pre- and post-panning DNA populations. The type of analysis based on a visual inspection of presence or absence of signals obtained from DNA spots that were characteristic for phage enrichment applied so far, only could be applied to a limited set of genes. This way, however, is not feasible for comparison of intensities of thousands of genes arrayed on the amygdala microarray. Two software packages provided by BioDiscovery (ImaGene 5.5.2 and GeneSight 3.5.2) were utilized for microarray image and data analysis, respectively. Following the fluorescence measurements made for Cy5 dye at each spot on the array, a local spot background subtraction and data normalization, the mean fluorescent signal values were acquired and presented in tables. Such data sets were subsequently analyzed using the Microsoft Excel software. Values obtained from both pre-panning (reference) and post-panning (experiment) microarrays were compared and signal ratios (experiment/reference) were calculated. The highest ratio values indicated an over-representation of post-panning genes that belong to phages enriched during affinity selection process. First, a careful enrichment analysis of the model phages on their peptide-ligands was performed utilizing both, visual and software driven comparison of hybridization signals on the microarray biochips. A standard experiment was carried out using a mixture containing the three model phages that were supplemented into the T7 human brain phage library (1.6×10^{10} pfu/ml) to achieve their final concentration of 1×10^{4} pfu (dilution ratio $1:1.6 \times 10^{6}$). Phages were selected on the 25 aa long peptide ligands for the three model domains (pep: E, F, Y). DNA populations isolated from pre- and post-panning phage mixtures after propagation were labeled with Cy5 dye and separately hybridized to the biochips. Control Fas DNA, supplemented to all mixtures was shown to hybridize to its complementary DNA on all biochips utilized in this experiment. A peptide specific enrichment of phages displaying a particular model protein domain was visible as a red fluorescence signal on a post-panning biochip, emitted from DNA-spot which code to the displayed protein. Absence of that signal on a pre-panning biochip indicated an under-representation of that particular domain within pre-panning phage population. Microarray analysis performed after panning on the pep-E (Fig 3.5.19), revealed that its DNA was over-represented within the post-panning population. A visual comparison of sections of the reference and experimental arrays, as well as signal intensity values measured using the software package for the three genes of interest (probes E, Y, F) are shown. Similarly, enrichment of phages displaying the WW/YAP domain on its peptide-ligand (pep-Y) was well visible on the corresponding biochip (Fig 3.5.20).

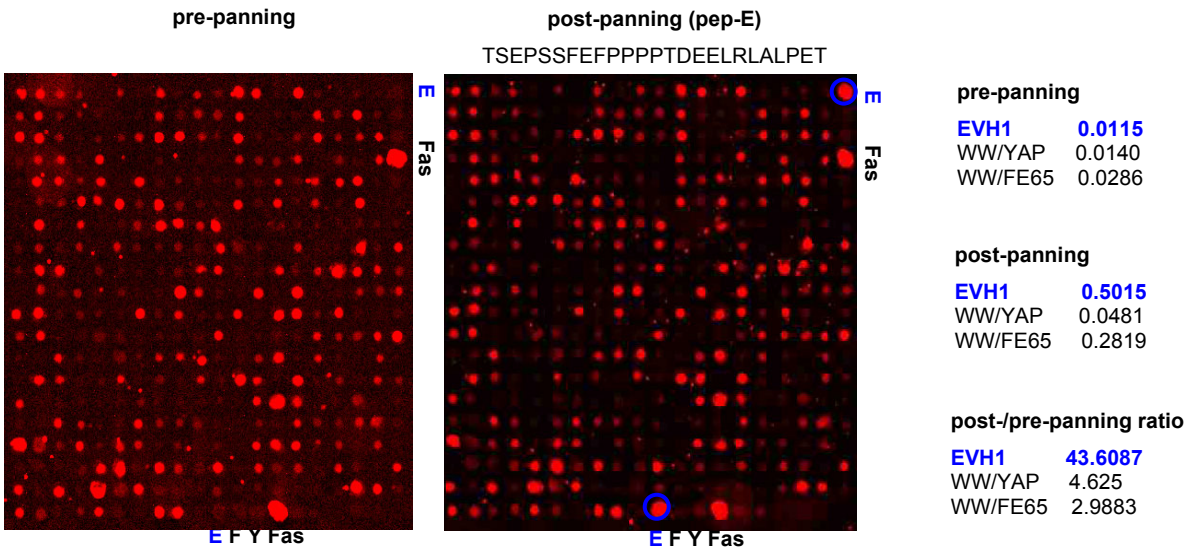


Fig 3.5.19: Sections of amygdala biochips showing hybridizations with Cy5 labeled pre- and post-panning DNA populations performed after phage selection over the EVH1 peptide-ligand (pep-E). The model phages were diluted with human brain T7 phage library at ratio 1:1.6 x 10^6. Blue circles and fonts indicate the gene coding for the EVH1 protein domain specific for the E target-peptide. Signal intensities from pre- and post-panning biochips and calculated rations are shown. Short name code: EVH1 (E), WW/FE65 (F); WW/YAP (Y); Fas antigen (Fas).

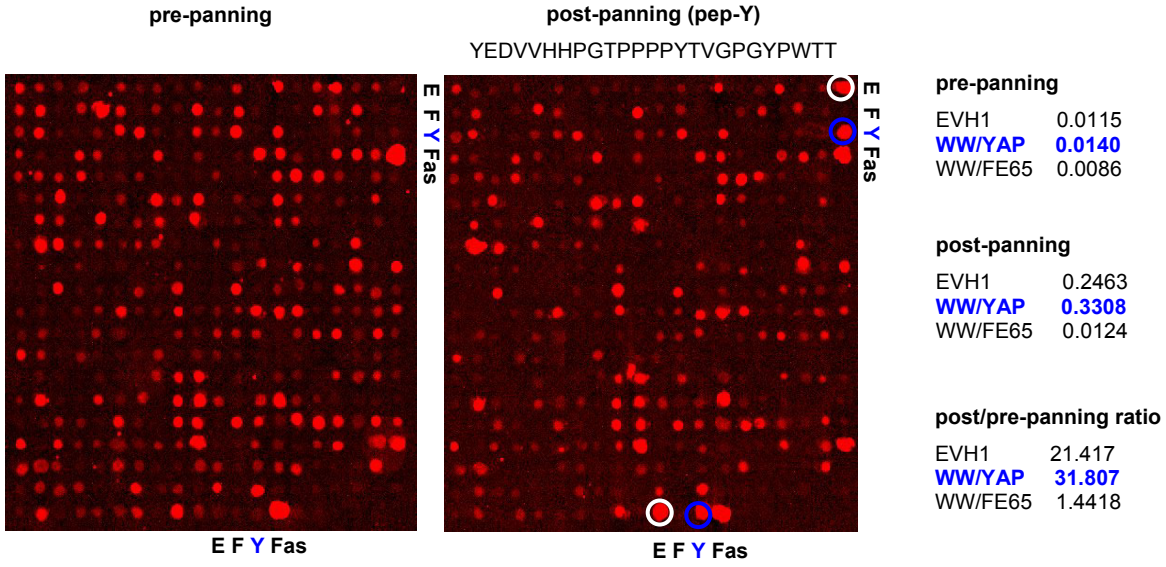


Fig 3.5.20: Sections of amygdala biochips showing hybridizations with Cy5 labeled pre- and post-panning DNA populations performed after phage selection over the WW/YAP peptide-ligand (pep-Y). The model phages were diluted with human brain T7 phage library at ratio 1:1.6 x 10^6. Blue circles and fonts indicate gene coding for the WW/YAP protein domain specific for the Y target-peptide. Signal intensities from pre- and post-panning biochips and calculated rations are shown. Short name code: EVH1 (E), WW/FE65 (F); WW/YAP (Y); Fas antigen (Fas).

Phages displaying the WW/FE65 protein domain were not detected within the population eluted from the pep-F ligand (Fig 3.5.21). This is in agreement with our previous panning experiments (see 3.2.1), which also showed that the enrichment of phages displaying this domain was always less efficient than the enrichment of phages displaying the other two model domains, because of lower affinity/avidity (see Bialek *et al.,* 2003). Unexpectedly, an additional signal obtained from the DNA probe E, was observed on both post-panning biochips corresponding to the pep-Y and pep-F ligands. Analysis of signal intensities by the software package revealed that this signal was only two times weaker in both cases than that obtained from the biochip related to target peptide pep-E. This issue will be discussed later in this section (3.5.7). Nevertheless, the successful identification of a specific enrichment of highly diluted phages (1:10^6) displaying the EVH1 and WW/YAP domains on their respective peptides-ligands, has proven that the microarray analysis takes an advantage over other approaches described previously, which failed to detect such enrichment. Therefore the microarray analysis should be considered as a very sensitive and promising method.

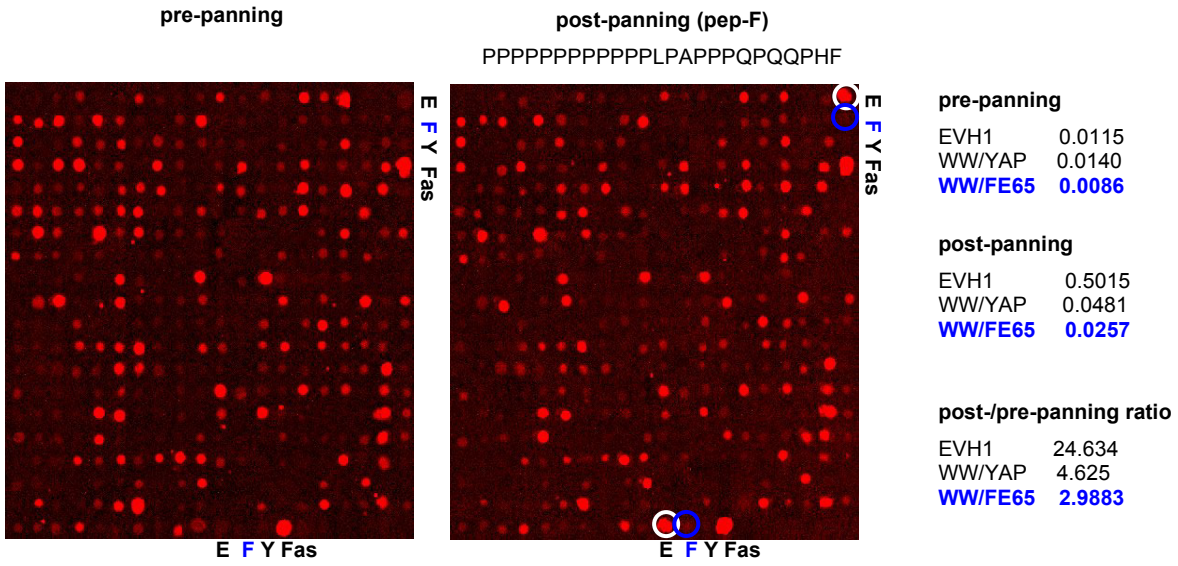


Fig 3.5.21: Sections of amygdala biochips showing hybridizations with Cy5 labeled pre- and post-panning DNA populations performed after phage selection over the WW/FE65 peptide-ligand (pep-F). The model phages were diluted with human brain T7 phage library at ratio 1:1.6 x 10^6. Blue circles and fonts indicate gene coding for the WW/FE65 protein domain specific for the F target-peptide. Signal intensities from pre- and post-panning biochips and calculated rations are shown. Short name code: EVH1 (E), WW/FE65 (F); WW/YAP (Y); Fas antigen (Fas).

An additional panning was carried out using the 25mer peptide-ligands pep-Y and pep-F to look into more detail to their more critical behavior in phage enrichments. The corresponding model phages were less than previously diluted within the total phage library (ratio 1:1.6 x 10^5). This additional experiment was expected to explain whether an increased concentration of possibly weaker binding WW/FE65 phages could promote a detectable enrichment on the corresponding target peptide. It could also be helpful in corroborating results observed in the previous screen, concerning the unexpected behavior of phages displaying EVH1 domain. The image analyses presented in Figs. 3.5.22 and 3.5.23 revealed a successful peptide-specific enrichment of WW/FE65 as well as WW/YAP model phages that was well visible on the applied biochips. The binding of EVH1 phage to both target peptides as observed before, was also detected in this case. These results support the former observations that the interaction of the WW/FE65 domain with its peptide-ligand must be weaker than the interaction of the WW/YAP and EVH1 domains with their targets. A weaker interaction can be compensated by a higher concentration of binding partners in order to detect enrichment, as observed in the case of the WW/FE65. This is also supported by enrichment of the WW/YAP domain, which was well visible in both above experiments, although the concentration of phages displaying this domain was the same as the concentration of phages displaying the WW/FE65.

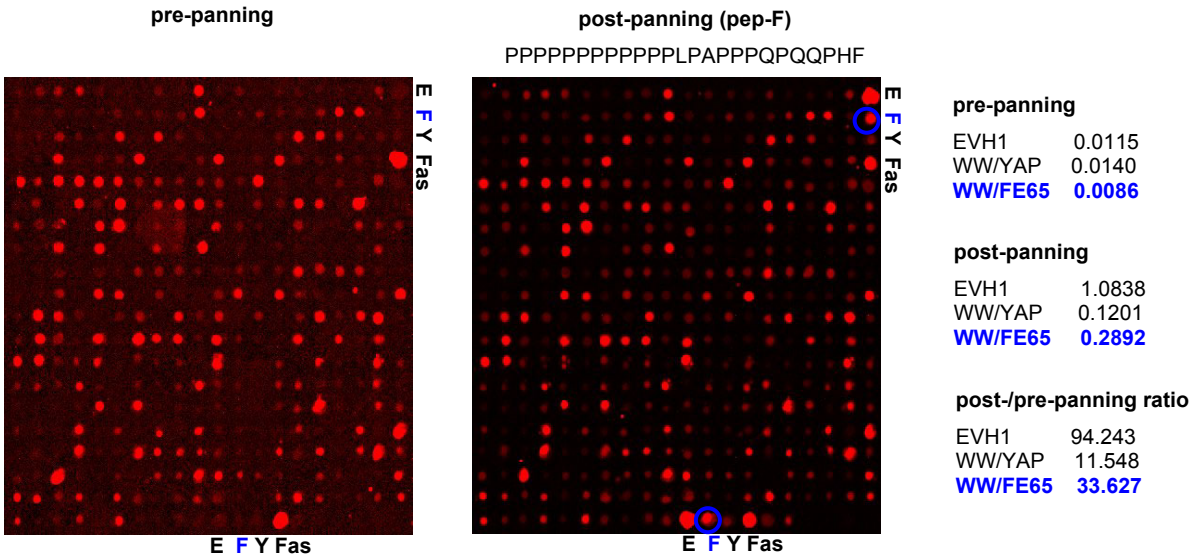


Fig 3.5.22: Sections of amygdala biochips showing hybridizations with Cy5 labeled pre- and post-panning DNA populations performed after phage selection over the WW/FE65 peptide-ligand (pep-F). The model phages were diluted with human brain T7 phage library at ratio 1:1.6 x 10^5. Blue circles and fonts indicate gene coding for the WW/FE65 protein domain specific for the F target-peptide. Signal intensities from pre- and post-panning biochips and calculated rations are shown. Short name code: EVH1 (E), WW/FE65 (F); WW/YAP (Y); Fas antigen (Fas).

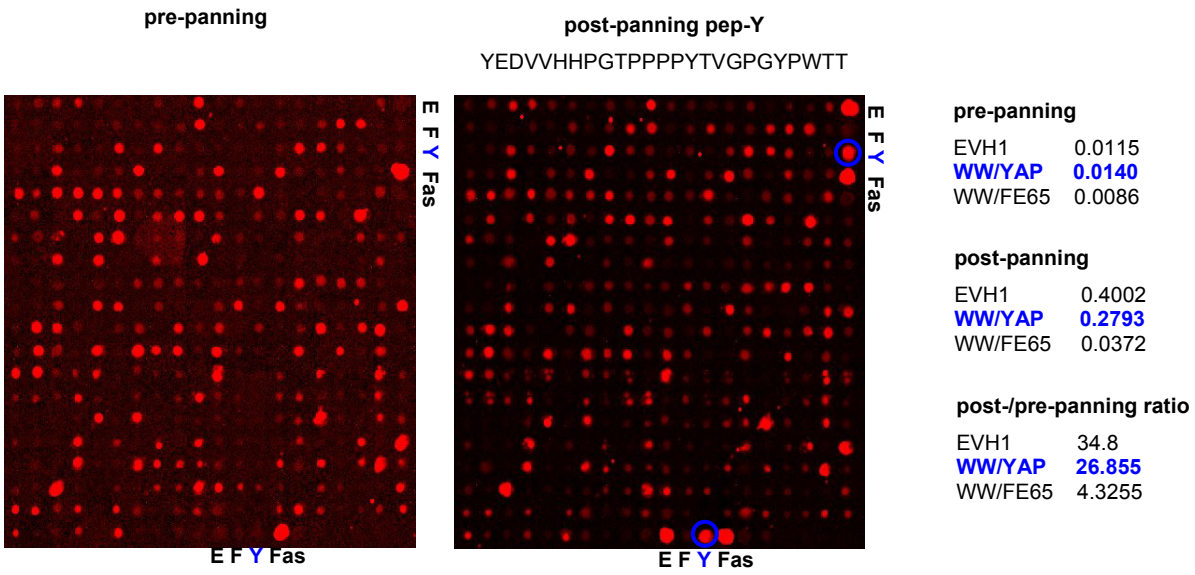


Fig 3.5.23: Sections of amygdala biochips showing hybridizations with Cy5 labeled pre- and post-panning DNA populations performed after phage selection over the WW/YAP peptide-ligand (pep-Y). The model phages were diluted with human brain T7 phage library at ratio 1:1.6 x 10^5. Blue circles and fonts indicate gene coding for the WW/YAP protein domain specific for the Y target-peptide. Signal intensities from pre- and post-panning biochips and calculated rations are shown. Short name code: EVH1 (E), WW/FE65 (F); WW/YAP (Y); Fas antigen (Fas).

In particular, both experiments indicated a minimum concentration for the WW/FE65 phages within a screened library (~10^5pfu, dilution ratio in this case = 1:10^5) that is required for a sufficient enrichment after only one round of affinity selection. Although the value of binding constant for the WW/FE65 protein domain is not determined, all observations suggest that its interaction with the peptide-ligand is weaker than interactions of the two other model domains with their respective peptides. According to the binding constant values, the EVH1 (5 μM) is more affine to its peptide than the WW/YAP (50 μM) domain to its corresponding peptide. This could be corroborated by the microarray analysis, as the values of signal intensities and pre-/post-panning ratios given by the EVH1 probes (Fig 3.5.19) were higher than those achieved from the WW/YAP probes (Fig 3.5.20). It is expected that enrichment of many proteins showing WW/FE65 like properties will be difficult to detect even using the sensitive array methodology. Many protein domains display minute and weak interactions with their targets and may be not detectable by this or any other methods in a one step panning process. A weak affinity of some proteins must be compensated with higher valence in order to increase the avidity and therefore enrichment efficiencies. Lower valence is preferable for enrichment of strong binding proteins. Therefore, a parallel utilization of the T7 and lambda phage systems characterized by low and high display capacities respectively, would allow for detection of a larger range of interactions (see 3.1). The studies performed

previously with both systems and the three model domains showed satisfactory enrichment of the weaker binding WW/FE65 in the case of the lambda system. However, the high valence of this system caused difficulties with elution of phages displaying the other two domains due to their strong interaction with the target peptides. The results achieved using the T7 system were consistent with the results obtained in the above experiments.

3.5.7 Cross-reactivity of the EVH1 domain

An explanation for the appearance of EVH1 signals indicating enrichment on both non-EVH1 ligand peptides stays unclear. There was no literature found, describing an interaction of the EVH1 domain from the murine protein Mena, with peptide-ligands for WW domains. Additionally, it was demonstrated that a protein from the same family, VASP (vasodilator-stimulated phosphoprotein) containing an EVH1 domain, didn't interact with peptides that are ligands for the WW/YAP (15 aa) and WW/FE65 (10 aa), (Niebuhr *et at.,* 1997). Therefore, specific interaction of the EVH1 from Mena with both 25 aa peptide ligands for WW domains utilized in these studies seems rather unlikely. Since the literature data exclude cross-interaction of both WW peptides with the EVH1 domain, another explanation for the puzzling observations from the microarray experiments was needed. The first and simplest one would be a too high concentration of the EVH1 phages that were supplemented to the phage population caused by mistakes in their titration or the preparation of the panning mixture. The over-represented phages would stick to the cellulose membrane and are detected as background. However, this hypothesis seems to be unlikely as the different model panning experiments were carried out independently, using separately prepared mixtures and freshly prepared phage lysates that were tittered separately, which minimized mistake probability. Additionally, this hypothesis couldn't be supported by microarray results achieved previously by Bialek (Bialek *et al.,* 2003), where representation of the EVH1 and WW/FE65 phages supplemented to the library was one and two orders of magnitude higher (f.c 1 x 10^6 pfu) than their representation in the experiments described here. This experiment showed that, despite a high representation of model phages, the enrichment of EVH1 was not observed on the pep-F, suggesting different nature of this problem. Another hypothesis that could possibly explain the observed behavior of the EVH1 domain assumes the presence of a DNA insert within the post-panning population, which is partially homologous to the DNA sequence of this domain. In order to be enriched, the hypothetical phage containing such insert must bind non-specifically to the pep-F and pep-Y spots, as there is no functional EVH1 domain displayed on its surface. The cross-hybridization of this insert to the EVH1-DNA would be visible as a positive signal on the array, therefore considered as phage enrichment. Such enrichment signal must be visible on the biochips corresponding to all

peptides in the case if the unspecific phage could bind to the fibers of the cellulose membrane, or aggregates with the coat proteins of other phages or with other displayed proteins. To check this hypothesis, an additional panning of the human brain phage library without a supplement of model phages was carried out on 25mer pep-E and pep-Y ligands. In the case of presence of the hypothetical phage containing cross-hybridizing insert, an enrichment signal given by the EVH1 DNA spot, would be visible after both pannings. However, the post-/pre-panning ratios calculated after microarray analysis, showed that there was no enrichment of the EVH1-like phages neither on its target peptide nor on the pep-Y (see Table 3.5.1). Ratios obtained after panning with the phage library that was supplemented with the model phages (f.c 1 x 10^4 pfu, see above), were ten-fold higher than ratios from the present experiment. Since Mena is a murine protein, its EVH1 domain was not expected to be present within the human brain phage library. Additionally, any homologous DNA sequences hybridizing to the EVH1 probe on the array have been found within the screened mixture, eliminating the cross-hybridization hypothesis discussed before. As already mentioned above, the literature records exclude a specific interaction of the EVH1 domain with peptide-ligands for both WW domains. Therefore the observed enrichment of this multi-valent phage on both proline-rich target peptides for WW domains was considered as unspecific.

Table 3.5.1: Comparison of fluorescent signal intensities from EVH1 DNA spot on the biochips corresponding to pep-E and pep-Y ligands, after affinity selection of T7 human brain phage library performed without and with supplement of the model phages

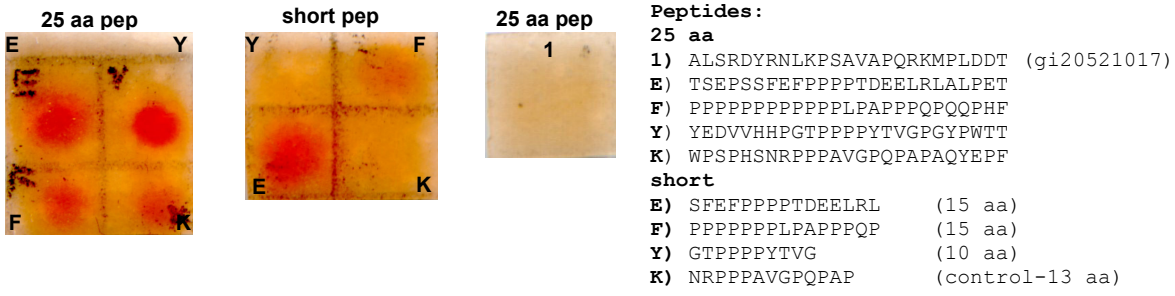
Peptide / Phage mix	Pre-panning	Post-panning	Post-/pre-panning ratio
pep-E/ (no supplement)	0.0115	0.0538	4.6782
pep-Y/ (no supplement)	0.0115	0.0343	2.9826
pep-E/ (supplement 1x10^4)[a)]	0.0115	0.5015	43.6087
pep-Y / (supplement 1x10^4)[a)]	0.0115	0.2463	21.417

[a)] supplement of the EVH1, WW/YAP and WW/FE65 model phages to the T7 phage library

In order to check for cross-binding of the EVH1 domain to the pep-Y and pep-F as well as to other indifferent peptides without phage context, a set of simple assays was performed utilizing the GST-MenaEVH1 fusion protein. Spots containing cellulose-bound peptides of different length (10-15 aa and 25aa), were incubated with the GST fusions, followed by incubation with anti-EVH1 antibody and subsequently incubated with a secondary antibody conjugated to peroxidase enzyme. Binding was detected by the enzymatic reaction with its substrate, as described in section 2.2.4.6. The first two assays were performed using different dilutions of primary antibody. Results obtained from both tests showed that in the

case of 25aa peptides, EVH1 domain bound to its target peptide (pep-E), but also to the poly-proline peptides for both model WW domains (pep-Y, pep-F), as well as to the control peptide (pep-K) containing 6 proline residues (Fig 3.5.24). Binding to another randomly chosen control peptide (1), with only three dispersed proline residues was not observed. Contrary, in the case of the short peptides, binding with the WW/YAP and the control peptide was not observed. Only a weak signal was detected on the pep-F spot. A significant difference in signal intensities between the spots pep-E and pep-F was well visible in the assay 1, with higher concentration of the GST-MenaEVH1 and lower dilution of the primary antibody (Fig 3.5.24). To exclude a possible unspecific binding of the protein to cellulose (trapping) that might be caused by the high concentration of the fusion protein and primary antibody used for the assay, concentrations of both proteins were reduced in the second assay. Signals obtained from all peptides were weaker in general when compared with signals from the first assay. Nevertheless, the above results were confirmed. The difference in signal intensities between pep-E and pep-F was however not as clearly visible as in the assay 1. It is possible that the concentration of fusion protein was to low in this case and therefore less EVH1 fusion protein was bound to the pep-E spot, resulting in weaker signal than observed above. This could also be caused by possible differences in handling of membranes. An additional, stronger signal given by the control peptide 2 was observed. This could be explained by possible strong non-specific affinity of this peptide to the GST-tag or entire fusion protein. The other control peptides, including proline rich peptide 6, didn't indicate any binding (Fig 3.5.24).

1. 2 µg GST-EVH1 + anti-Mena PAK 73 Ab (dil. 1:1000) + POD anti-rabbit IgG (1:2000)



2. 1 µg GST-EVH1 + anti-Mena PAK 73 Ab (dil. 1:3000) + POD anti-rabbit IgG (1:2000)

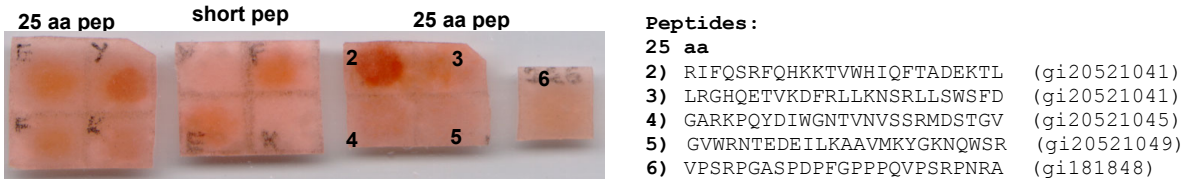
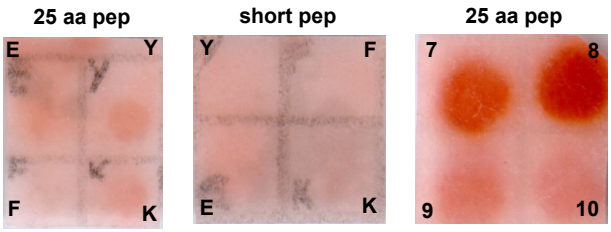


Fig 3.5.24: Binding assays of the GST-MenaEVH1 with various peptide ligands of different length.

Signals obtained from both 25mers (pep-F, pep-Y) in both assays, suggested interaction the EVH1 fusion protein with these peptide-spots. These results are similar to the results that were obtained after panning experiments on the same peptides (see 3.5.6). Analogously, the enrichment of the EVH1 phages was detected on both biochips corresponding to both non-EVH1 peptides. In order to check for a possible non-specific interaction of the GST-tag protein with the model peptide-spots, an additional test was performed with this protein (1μg) alone. Detection was carried out using rabbit anti-GST antibody (1:2000) with conjugated peroxidase enzyme. Interactions were observed for all four control spots (7-10) with strong differences in signals intensities (Fig 3.5.25). No signals indicating interaction of the GST-tag alone with the short model peptides were detected on these spots. In contrast to this, as described before for the assays 1 and 2, the short model peptide (pep-F) showed signal indicating interaction with the GST-EVH1 fusion protein (Fig 3.5.24). This comparison suggested that the interaction with this peptide must be mediated by the EVH1 alone, not by the GST protein. Moreover, interaction of this fusion-tag with the 25mer pep-F was not detected as well (Fig 3.5.25). The signals from this peptide were however visible when the GST-EVH1 fusion was used (Fig 3.5.24). This suggests again that the observed cross-interaction was mediated by the EVH1 domain. Additionally, significant signals indicating binding of the GST-tag alone to the 25mer pep-E peptide were not observed; relatively weak signals were visible on the 25mer pep-Y and control peptides (Fig 3.5.25), demonstrating a contribution of the GST-tag in interactions with these peptides. Unfortunately, it was not possible on the basis of these experiments to completely prove whether the interaction is driven be the GST or the EVH1 itself. Additionally one could suspect that it is driven by both these proteins, as the previous experiments utilizing microarrays, where no GST-tag was used, showed similar non-specificity.



Peptides:

25 aa

E) TSEPSSFEFPPPPTDEELRLALPET
F) PPPPPPPPPPPPLPAPPPQPQQPHF
Y) YEDVVHHPGTPPPPYTVGPGYPWTT
K) WPSPHSNRPPPAVGPQPAPAQYEPF
7) PQPMEQGSSGVLGGKEEGGKPERSS (gi20521013)
8) VTLVEELSLQEAMRKEPGESSSRKA (gi20521013)
9) QKAFDNTEGKREGYSADSAQEASPA (gi20521051)
10) ADSAQEASPARQNVSSASNPENDSS (gi20521051)

short

E) SFEFPPPPTDEELRL (15 aa)
F) PPPPPPPLPAPPPQP (15 aa)
Y) GTPPPPYTVG (10 aa)
K) NRPPPAVGPQPAP (control 13 aa)

Fig 3.5.25: Binding assays of the GST (1μg) protein with various peptide ligands of different length.

Nevertheless, based on the literature data (Niebuhr *et al.,* 1997; Prehoda *et al*., 1999; Carl *et al*., 1999; Ball *et.al.,* 2002), showing very high specificity of EVH1 domain to the FPPPP core motif flanked by acidic residues, a specific binding of the EVH1 to the 25 aa peptide-ligands of WW domains seems rather unlikely. The results obtained here for the EVH1 domain and the results achieved previously for the EVH1 phages using microarrays support a higher non-specificity of this domain to the 25mer peptides compared to the shorter peptides. Nevertheless, comparison of interactions mediated by a single EVH1 domain with the multivalent phages might be inappropriate and therefore several additional, independent tests must be carried out in order to draw appropriate conclusions concerning these observations.

3.5.8 Statistical evaluation of microarray results

All results presented in this section so far, were obtained from the three protein domains, which were used as model proteins in order to test performance of the new microarray approach. Since the “home” produced biochips contain a total representation of 6144 different genes originating from amygdala region of human brain, a simultaneous detection of potential binders from human brain library that were enriched on peptide-spots became possible. Experiments performed using model phages revealed that the microarray profiling was more sensitive and efficient than the other applied systems. Following analysis with the gene expression profiling software suite, data filtration and statistical analysis were carried out using Microsoft Excel software, in order to select possibly most specific binding candidates. Previous microarray analysis performed for the three model domains selected on their corresponding peptides, was useful in establishing threshold values and filtration methods, which will be uniformly used for analysis of microarray data. Filtration of data sets will be achieved by elimination of weak and non readable fluorescent signals due to background and other technical problems, as well as those signals, whose intensities are equal on both pre- and post-panning microarrays. This process would allow for elimination of genes which correspond to phages that were not enriched on peptide-spots and genes corresponding to phages equally represented in both pre- and post-panning populations. Three important data sets are acquired after each microarray analysis. First data set gives mean values of Cy5 signal intensities measured for two replicas of each gene that are present on the amygdala biochip. The values for EVH1, WW/YAP and WW/FE65 domains acquired after analysis of post-panning biochips, ranged from 0.2 to 0.5. Therefore, the threshold value chosen for future analysis was set to larger as or equal 0.1, meaning that all values which are lower than 0.1 will not be taken into account. Second set, gives coefficients of variance (CV) for two replica gene spots on the biochip. This selection parameter was

applied for both pre- and post-panning biochips in order to exclude those spots, which are characterized by large CV values. Large CV values are obtained when one of the two replica spots is covered by contaminating background or dust particle, which give high fluorescent signal. Since fluorescence intensities of these two spots are usually not even similar (low correlation), their CV value tend to be very high. The CV values found for the model genes ranged from 0.1 to 0.3. Therefore, a general CV threshold level was set to 0.5. The third data set contains ratios calculated by dividing mean values of intensities of post-panning signals through pre-panning signals. Ratios are calculated for each probe (gene) on the microarray. The ratio values indicate phage enrichment on certain peptide. A high value of a post-panning signal emitted from a certain probe on the array divided by a low value of a pre-panning signal read out from the same probe gives a high ratio, which means positive phage enrichment. The threshold for this data set was also established from ratios obtained in the model experiments; EVH1 (43.608), WW/YAP (31.801), WW/FE65 (24.634). A threshold filter was chosen that preserved only those gene probes on the array, which were characterized by ratio of pre-/post-panning signals equal or higher than 15. A list of data explaining filtration process using the values obtained from model panning over the pep-Y is shown in Table 3.5.2. Data corresponding to only those genes, for which all the filtration requirements were fulfilled (blue + green+ red) will be taken into account in the future analysis.

Table 3.5.2: Example showing filtration of microarray data acquired after panning on pep-Y, using three experimentally defined threshold values. post-panning spot signal ≥ 0.1, post and pre-panning CV ≤ 0.5, ratio ≥ 15

Data before filtration

	Spot signal		CV value		Ratio
Gene ID	Pre-panning	Post-panning	Pre-panning	Post-panning	Post-/Pre
DKFZp761P241Q	0.1753	0.2274	0.1786	0.0092	1.297204792
DKFZp761P2420Q	0.0291	0.0241	0.0909	0.2076	0.828178694
DKFZp761P242Q	0.0538	0.0876	0.0747	0.1029	1.628252788
DKFZp761P248Q	0.0154	0.0282	0.3441	0.146	1.831168831
E	0.0115	0.2463	0.2454	0.2022	21.4173913
Ezr	0.0063	0.0072	0.5727	0.4483	1.142857143
F	0.0086	0.0124	0.3162	0.354	1.441860465
Fas	1	1	0.1899	0.3179	1
IL-4	0.0088	0.0059	0.4479	0.7667	0.670454545
Y	0.0104	0.3308	0.3863	0.2491	31.80769231
Empty spot (buffer)	0.0012	0.0047	0.9378	0.7106	3.916666667

Data after filtration

	Spot signal		CV value		Ratio
Gene ID	Pre-panning	Post-panning	Pre-panning	Post-panning	Post-/Pre
E	0.0115	0.2463	0.2454	0.2022	21.4173913
Y	0.0104	0.3308	0.3863	0.2491	31.80769231

The above filtration process has been demonstrated to be useful for the microarray analysis of protein interaction data from one enrichment experiment on a peptide-spot. However, further statistical analysis is necessary to prove whether a selected protein is a potential specific binder. At least two of the same but separate panning “sister” experiments have to be performed for each peptide-ligand to provide data for a minimal statistical analysis. In order to be included in a pool of potentially specific interaction partners, a phage which was selected in one panning experiment must be also selected in the “sister” experiment and can not be abundantly appearing in other panning experiments performed on different peptide-ligands. A phage, which is selected in one “sister” experiment only, will be considered as a random background. Other reasons for differences in gene representation between two the same data sets could be (i) a strong enrichment of unspecific phages in one panning experiment or (ii) possible differences in PCR amplification resulting in over- or under-representation of genes. According to the rules established for “sister” biochip comparison, these candidates which occurred only in one data set have to be eliminated.

If enrichment of this phage is observed in all performed experiments, this means that such candidate should be treated as constant background. It is likely that some phages will bind to one peptide-ligand only but in unspecific manner (false positive). Such phages will not be deselected and, thus, identified as specific binders. Therefore any result from the microarray analysis need to be verified by other independent methods. The discrimination of potentially specific and unspecific candidates, based on statistical analysis is demonstrated below, using microarray data acquired for the three model phages after affinity selection on their target peptides (Table 3.5.3).

Table 3.5.3: Example showing identification of specifically and not specifically enriched phages, using comparison of statistical data from post-/pre-panning ratios obtained from "sister" and unrelated panning experiments. Post-/pre-panning ratios showing: specific enrichment of WW/YAP domain on its peptide (red), leak of enrichment of WW/YAP domain on unrelated peptide-ligands (green), specific enrichment of EVH1 domain on its peptide-ligand (orange) and unspecific enrichment on unrelated peptides (blue). Filtration parameters: signal ≥ 0.1; CV ≤ 0.5; ratio ≥ 15.

Domain	Post-/pre-panning ratio				Judged enrichment
	pep-Y		pep-F	pep- E	
	Experiment 1	Experiment 2			
EVH1	21.4173	34.8	94.2434	43.6087	non specific
WW/YAP	31.8076	26.8557	11.5480	4.625	specific
WW/FE65 (1)	1.4418	4.3255	33.6279	3.3255	specific

(1) Analysis based on panning experiment with supplement model phage (1×10^5)

Enrichment of the WW/YAP domain was detected in both panning experiments on the corresponding peptide ligand (red) but was not detected in the other two experiments on different peptides (green), strongly indicating a specific domain-ligand interaction. As discussed above, most probably due to a weak interaction with its peptide, the enrichment of the WW/FE65 domain was not observed after panning with a library containing 1×10^4 pfu (dilution ratio $1{:}10^6$) of model phages but was detected after supplementing with the $1{:}10^5$ pfu (dilution ratio $1{:}10^5$). Because there was only one experiment performed using this higher concentration of model phages, the statistical comparison with "sister" experiment cannot apply in this case. However, the fact that this domain was enriched only on its corresponding peptide proves a specific binding. Enrichment of the EVH1 domain was observed in all four experiments and also in pannings on non-EVH1-peptides (blue). According to the rules established, results obtained for EVH1 should be eliminated from the pool of specifically binding candidates. This is in agreement with the previously discussed observations of high non-specificity of this domain for the 25mer peptides used in this work. The judgment on the EVH1 domain seems to be proper in this case, although its reported highly specific interaction with peptides containing the FPPPP motif. It also suggests that the statistical analysis is very sensitive and restrictive and could be advantageous during analysis of unknown peptide interacting proteins.

3.5.9 Phage library expansion by infection and propagation in bacteria

As already discussed in the section 3.3, panning experiments on peptide arrays require large volumes of phage library lysates. Since the volume aliquot of the commercially available T7 human brain library doesn't exceed 100 µl, an amplification step is necessary before each screening. Additionally, each post-panning phage eluate needs to be propagated in order to provide enough material for PCR amplification. This is done by infection of host bacteria with the phages from the original library and further recovery of resulting phage lysate (see 2.2.4.2). It is absolutely required, that this step must not change the relative abundances of the phages in the library. Otherwise, faster growing phages will be identified as false positives. In order to test which phages tend to amplify faster and are therefore over-represented within the library after its propagation, a simple comparison experiment was carried out using the amygdala microarrays. First, an aliquot of 10 µl of the original T7 library was used to infect 20 ml of host bacteria and phages were propagated. Another 10 µl aliquot of this new phage library was used for a second propagation in the same way. Since the only difference between these two libraries is the additional phage amplification step, they were called "amplified 1°" and "amplified 2°". Next, an aliquot of 100 µl from each library was used for isolation of phage DNA, which was subsequently used for a direct labeling with the Cy5-dye during PCR. The same amount of DNA corresponding to each library was hybridized to two separate amygdala biochips. Microarrays were analyzed using previously mentioned software suite. Signal intensities emitted from of gene probes present on both arrays were compared with each other. Correlation coefficient value for both libraries was 89 % (r^2 = 0.8927). Genes showing similar signal values and therefore indicating similar representation of phages for both experiments were preserved. Outstanding genes showing large differences in intensities of emitted signal were excluded from the array list, which will be subsequently used for all analyses of phage enrichments. The fold change (FC) thresholds applied for data cleansing, were set greater than -1.5 and less than 1.5. Fig 3.5.26 shows a scatter plot of signal intensities before (A) and after (B) data cleansing. Genes chosen as a uniform set for phage enrichment analysis, were located between y = 2x and y = 0.5 regression lines. In effect, the cleansing process increased correlation coefficient for these genes to 97 % (r^2 = 0.9744). According to these results, 33 % of genes printed on glass slides were shown to correspond to phages, which seem to propagate faster than other library members and most likely therefore contribute to unspecific background. The final number of 4098 genes from the array list was taken into account during further microarray analysis.

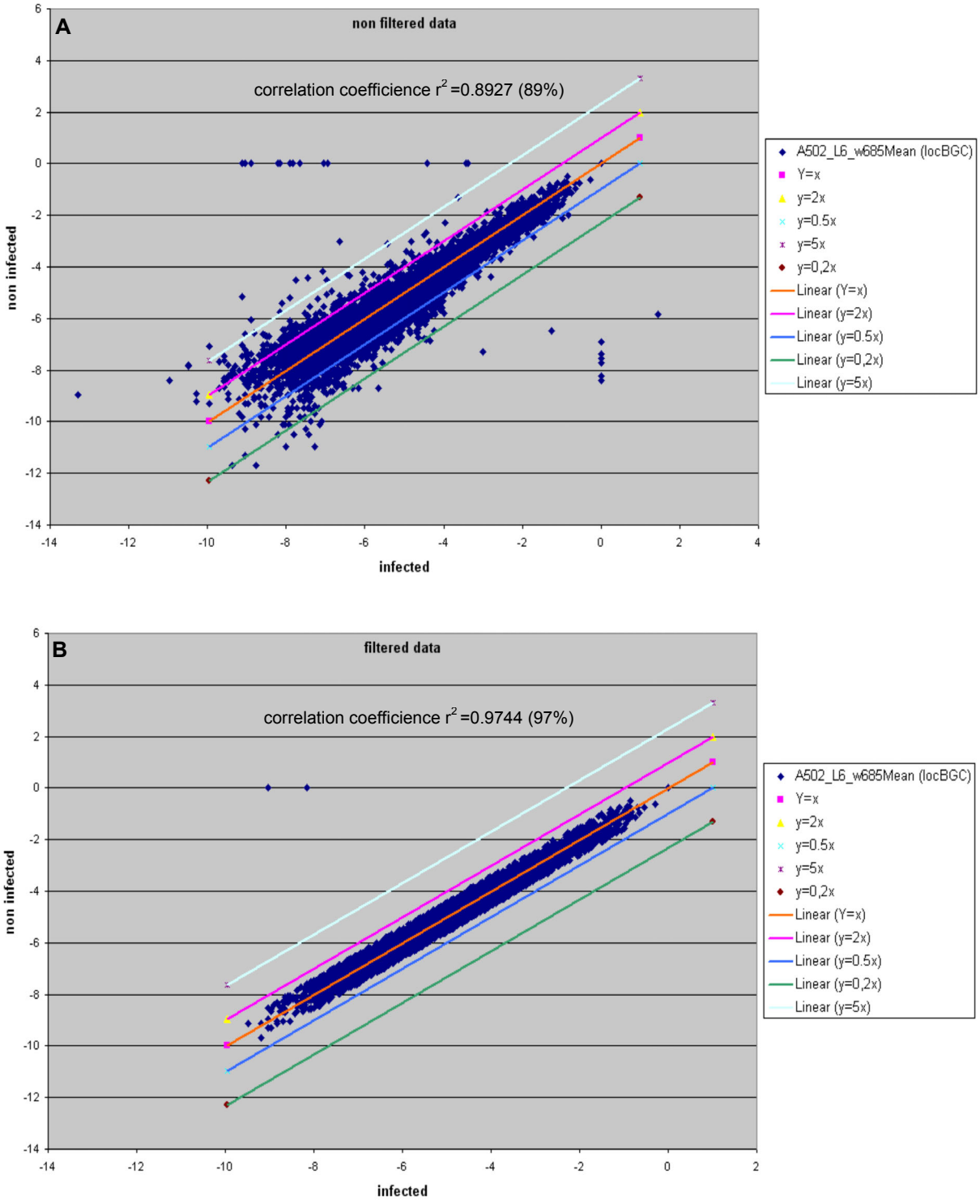


Fig 3.5.26: Scatter plots showing relationship between phage libraries which were propagated one time (amplified 1°) and two times (amplified 2°) before (A) and after (B) data cleansing. Points placed close to the regression line (y = x), represent genes with similar values for both experiments. Points, which are far away from this line have low correlation score and were excluded from the gene array list. All points within the two-fold borders (y = 2x and y = 0.5) were preserved. Five-fold borders are indicated by y = 5x and y = 0.2x respectively. A502 – ID of biochip related to non infected library.

3.5.10 Expanding the interaction screen to other domains and peptides

Previously, only the three model genes were taken into account for panning experiments and during microarray analysis in order to establish rules for data segregation and validation of phage enrichment. As the entire biochips were used for hybridization with post-panning DNA populations, the data achieved from other gene spots were subsequently used to verify possible enrichment of other phages on model peptides. In addition, several screens were carried out, using different peptides and the T7 library without supplement of model phages. Proline rich peptides were preferred as targets, as proline plays a special role in peptide mediated protein-protein interactions and is involved in interaction with many different protein domains (Kay *et al.,* 2000; Macias *et al.,* 2002; Sidhu *et al.,* 2003). Peptides were selected from native protein sequences using the ASS software tool (see section 3.2). Peptide sequences and database codes of their parental proteins are shown in (Table 3.5.4).

Table 3.5.4: Peptides used for phage affinity selection experiments. Peptides were selected using the ASS software tool.

Peptide ID	Peptide sequence	Protein ID
EVH1	TSEPSSFEFPPPPTDEELRLALPET	spP33379
WW/YAP	YEDVVHHPGTPPPPYTVGPGYPWTT	spP97764
WW/FE65	PPPPPPPPPPPPLPAPPPQPQQPHF	spQ03173
107	RLDLPPYPSYSMLYEKLLTAVEETS	gi20520999
113	TVWDKHSVTAATPPPSPTSGESGDL	gi20521017
148	TQQSSLPPPPPPDSLDQPQPMEQGS	gi20521013
199	PPLPPPPPLPKTPRGGKRKHKPQAP	gi20521051
200	GKRKHKPQAPAQPPQQSPPQQPLPQ	gi20521051

Each panning experiment was performed in duplicate. Microarray data were filtered using previously defined thresholds (signal ≥ 0.1; CV ≤ 0.5; ratio ≥ 15) followed by statistical analysis. It was shown that in some cases the number of genes was strongly reduced, already after a first filtration process. This was especially true for biochips with overall lower signal intensities, which could be caused either by higher background or lower quality of the hybridization process. Subsequent comparison of both, more and less reduced sets of genes frequently resulted in elimination of majority or sometimes all selected candidates. In the last step of analysis, an already reduced set of genes was subjected to another comparison with biochips, filtering those, which have tendency to occur in other panning experiments. The same genes selected for both “sister” experiments but also found in several other screens, were also excluded from list of potential candidates, as they likely correspond to non-

specifically binding phages. Candidates matching all requirements of statistical comparison were considered as potential peptide interacting partners. An example showing the analysis of data sets containing various numbers of genes obtained from three parallel pannings is described below. These pannings were carried out on the peptide-ligand for the WW/YAP domain using the T7 library with and without addition of model phages. The numbers of positively selected genes after the first filtration process are listed separately for the three experiments in (Table 3.5.5). Except for the two genes corresponding to the model domains, their enrichment was already discussed earlier; only one gene (DKFZp761O0120Q) was found in all three "sister" experiments. Very high values of post-/pre-panning ratio for this candidate suggested its specific interaction with the WW/YAP peptide.

Table 3.5.5: Three data sets obtained after initial filtration of microarray records for panning experiments performed on peptide for WW/YAP domain using phage library only and phage library supplemented with the three model phages. Filtration parameters: signal ≥ 0.1; CV ≤ 0.5; ratio ≥ 15. Genes, which positively passed all three types of selection, are showed in bold blue.

Exp 1. panning with supplement of model phages (1 x 10^4 pfu, dilution 1 : 10^6)					
Gene ID	**Signal intensitie**		**CV value**		**Ratio**
	pre-pan	post-pan	pre-pan	post-pan	post-/pre-pan
DKFZp761O0120Q	0.0032	0.2332	0.3834	0.2309	**72.875**
EVH1	0.0115	0.2463	0.2454	0.2022	**21.4173913**
WW/YAPY	0.0104	0.3308	0.3863	0.2491	**31.80769231**

Exp 2. panning with supplement of model phages (1 x 10^5 pfu, dilution 1 : 10^5)					
Gene ID	**Signal intensitie**		**CV value**		**Ratio**
	pre-pan	post-pan	pre-pan	post-pan	post-/pre-pan
DKFZp761C1516Q	0.0063	0.1024	0.127	0.0757	16.25396825
DKFZp761D2310Q	0.0053	0.1124	0.1534	0.3094	21.20754717
DKFZp761L2112Q	0.0054	0.1031	0.2239	0.247	19.09259259
DKFZp761L242Q	0.0071	0.1787	0.1015	0.1962	25.16901408
DKFZp761O0120Q	0.0032	0.2604	0.3834	0.1645	**81.375**
DKFZp761O0216Q	0.0056	0.1036	0.0123	0.1531	18.5
DKFZp761O2113Q	0.0067	0.1222	0.1153	0.4529	18.23880597
DKFZp761O2412Q	0.0055	0.1072	0.3061	0.0317	19.49090909
DKFZp761P198Q	0.0124	0.217	0.6201	0.127	17.5
EVH1	0.0115	0.4002	0.2454	0.1357	**34.8**
WW/YAP	0.0104	0.2793	0.3863	0.1315	**26.85576923**

Exp 3. panning with T7 phage library only					
Gene ID	**Signal intensitie**		**CV value**		**Ratio**
	pre-pan	post-pan	pre-pan	post-pan	post-/pre-pan
DKFZp761F1212Q	0.0305	1.4466		0.1956	47.4295082
DKFZp761F1510Q	0.0222	3.5795		0.1829	161.2387387
DKFZp761F1611Q	0.0109	0.4861		0.3487	44.59633028
DKFZp761O0120Q	0.0032	0.317		0.3834	**99.0625**

Comparison of data sets obtained for all experiments revealed that this gene was also over-represented in post-panning population, acquired after phage selection on peptide 107B (Table 3.5.6). Unfortunately, a comparison with a second “sister” biochip for this peptide was impossible due to hybridization failure. Nevertheless, high ratio value for this gene suggested a specific interaction of the related phage to this peptide.

Table 3.5.6: Statistical comparison of post-/pre-panning ratios obtained for gene DKFZp761O0120Q after all performed panning experiments.

DKFZp761O0120Q clone	
Peptides used for panning	**Post-/pre-panning ratio**
pep-Y exp 1	**72.875**
pep-Y exp 2	**81.375**
pep-Y exp 3	**99.0625**
pep-F	1.9375
EVH1	1.96875
200A	1.84375
200B	2.21875
148A	1.34375
148B	0.9375
199B	0.96875
pep113A	0.90625
113B	0.59375
107B	**65.09375**

Sequencing of the DKFZp761O0120Q gene followed by BLAST homology search revealed that it codes for a human Nedd-4-like E3 ubiquitin-protein ligase called also WW domain-containing protein 1 (WWP1) or Atropin-1 interacting protein 5 (AIP5). This protein contains four WW domains (class I), which bind (L/P)PxY motif (x-any amino acid). Interaction of WWP1 with the peptide-ligand for WW/YAP containing the PPPY motif was already identified previously using TAIS technology (section 3.4). Additionally, this protein was shown here as interaction partner for peptide 107. Analysis of the peptide sequence revealed that it contains the binding motif LPPY for WW domains of class I, (Kasanov *et al.,* 2001). Moreover, peptide 107 is a fragment of the Nedd 4-like ubiquitin ligase 1 (NEDL1) encoded by KIAA0322 gene and is located in its C-terminal HECT domain. A similar motif, PPYK, was identified in the C-terminus of WWP1 protein sequence (Fig 3.5.27). The fragment of the nucleotide sequence of DKFZp761O0120Q clone as well as amino acid sequences of both proteins are listed in appendix 4. Interaction of WWP1 with a peptide originating from NEDL1 and the presence of

a similar motif in WWP1 itself, suggests possible interaction of these two proteins with each other and likely with other proteins containing WW domains of class I. These speculations however, have to be additionally proven by other methods validating in vivo protein-protein interactions. It is also not known, whether this finding could be of biological relevance. This can now be investigated using the peptide in competition experiments or by mutating the corresponding site in the protein by site directed mutagenesis and inspection of the biological consequence. Furthermore, this discovery proves also a proper performance of the novel Accessible Surface Scanner software tool (ASS), which was applied for selection of the 107 peptide from the WWP1 protein sequence.

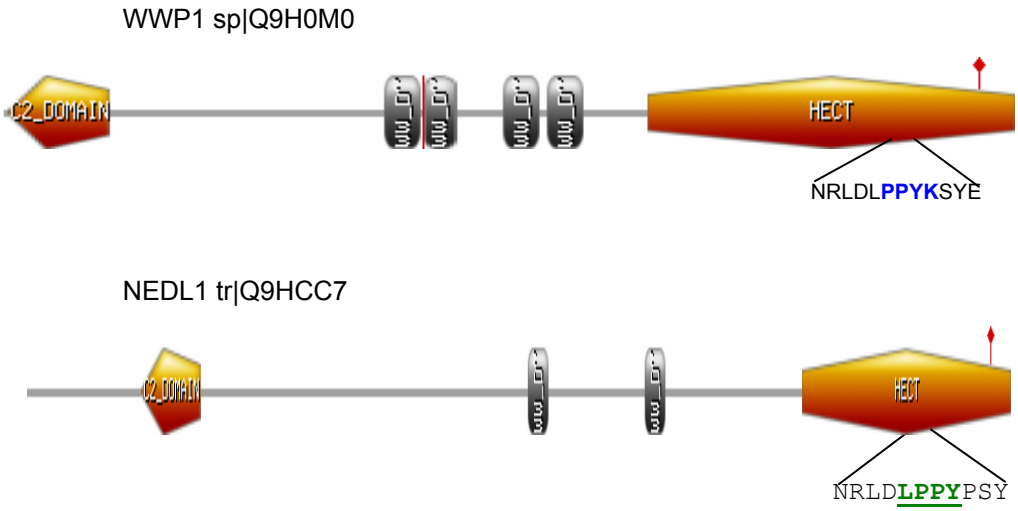


Fig 3.5.27: Graphical domain representation of WWP1 and NEDL1 proteins. Fragment of peptide 107 (green). Similar motif found at the C-terminus of WWP1 (blue).

The example above showed that the microarray analysis based on thresholds defined with the help of model system is quite restrictive. Apart from the positively enriched gene coding for the two described proteins, only one additional gene DKFZp761L0517Q passed the entire selection process after analysis of microarray data achieved from all panning experiments. Phage corresponding to the above clone was enriched on the ligand for the WW/FE65 domain (pep-F) in both selection experiments and was not detected in any other experiment. Nevertheless, sequencing of this gene revealed that it codes for myelin basic protein (MBP). Myelin basic protein is a hydrophilic protein that may function to maintain the correct structure of myelin, interacting with the lipids in the myelin membrane by electrostatic and hydrophobic interactions (Inouye and Kirschner, 1991). Since FE65 protein binds its proline rich peptide through WW domain, one would expect a presence of this domain within the

identified sequence. However, none of known domains interacting with proline rich peptides were found in this protein. Additionally, the MBP protein is not at all related to FE65. Analysis of post-/pre-panning ratios for this gene obtained in both “sister” experiments revealed that these were only a little higher than the defined threshold limit (15.4794 and 15.5616). This observation, together with lack of any relationship of this protein with the FE65 protein lead to the conclusion that phages presenting this protein were most probably enriched on the peptide-spot as difficult to recognize false positive background. An increase of the ratio threshold (i.e. >20) in this case would eliminate this candidate. Thus, a definition of uniform thresholds for global analysis seems to be very difficult due to different affinities of peptide specific proteins. The nucleotide sequence of the DKFZp761L0517Q clone as well as amino acid sequence of MBP protein is shown in appendix 4.

The fact that no other candidates that were enriched in different pannings were qualified as valid peptide-binding partners could suggest that the defined threshold values for filtration of microarray data and rules of statistical analysis are too stringent and may eliminate most of the binding candidates. On the other hand, comparisons of microarray data, especially post-/pre-panning ratios obtained from all above experiments, showed that ratios for specifically enriched proteins are usually relatively high, in contrast to those observed for the false positive like MBP (Table 3.5.7). A moderate increase of the defined thresholds, thus, should not eliminate specific binders.

Interestingly, comparison of post-/pre-panning ratios revealed certain relationship between the value and the affinity of the protein to its target peptide. Very high values observed for WWP1 protein are related to multiple WW domains present in this protein. Four binding modules within this protein may account for its high avidity, thus, more efficient enrichment on the target peptide. This was also observed previously using the TAIS technology. Very efficient enrichment of WWP1 protein was found in that case, whereas the single WW/YAP domain that was supplemented to the phage library at final concentration 1×10^4 pfu (dilution $1:10^6$), was not detected (see section 3.4). Assuming that the screening of very high diversity phage libraries has to cope with protein-peptide interactions of a wide range of binding properties, the maintenance of relatively tolerable thresholds is important, even though the risk of isolation of non specific interactions may become higher. In order to test whether application of lower thresholds than applied for the above microarray analyses, would result in the identification of a higher number of potentially specific interaction partners, all data were additionally filtered using another set of signal with CV and variable ratio values slightly decreased (signal ≥ 0.1; CV ≤ 0.5; ratio ≥ 5, ≥ 10). Although the representation of genes in all experiments became significantly increased, only a few candidates occurred in both “sister” experiments and these were usually eliminated when comparing with data obtained from other biochips (see appendix 5). These analyses indicate

that relaxing the threshold values will not result in an increased number of positive interaction candidates. But, it may increase the probability of selecting false positive background phages like MBP. This analysis supports again the fact that thresholds applied for data selection which were defined on the basis of real model interactions are selective enough to omit most unspecific interactions with preservation of potential specific binding candidates.

Table 3.5.7: Comparison of post-/pre-panning ratio values observed for specific interactions and those considered as non specific interaction (MBP).

DKFZp761O0120Q clone	
peptides / panning	**Post-/pre-panning ratio**
pep-Y / exp 1	**72.875**
pep-Y / exp 2	**81.375**
pep-Y / exp 3	**99.0625**
107B	**65.0937**
WW/YAP model protein	
pep-Y / exp 1 (f.c $1x10^5$)	**31.8076**
pep-Y / exp 2 (f.c $1x10^4$)	**26.8557**
EVH1 model protein	
pep-E / (f.c $1x10^4$) [a)]	**34.8**
FE65 model protein	
pep-F / (f.c $1x10^5$)	**33.627**
MBP (non specific)	
pep-F exp 1	**15.5616**
Pep-F exp 2	**15.4794**

[a)] Despite its non-specific interaction with the 25mer peptides (pep-F, pep-Y, pep-K), observed in these studies (section 3.5.7), the EVH1 domain was included in this table as its interaction with the pep-E ligand is highly specific (Niebuhr *et al.,* 1997)

3.5.11 Conclusions

Studies performed in this section for a selected number of target-peptides, confirmed a high potential of the microarray based approach for high-throughput performance of the entire process of a genome-wide mapping of peptide-mediated protein-protein interactions. Compared to the other approaches, more focused on certain types of peptide ligands (Salcini *et al.,* 1997; Sparks *et al.,* 1994; Tong *et al.,* 2002) this approach here is particularly able to discover new families of peptide-recognizing protein domains from genome wide screenings. Several important requirements have been identified which have to be met in order to perform successful screening experiments. These include, highly representative genome spanning repertoires of peptide fragments, phage-based cDNA expression systems as well as microarray biochips which are optimally matching the screened phage libraries. Peptides can be selected from native ORF sequences, as described in this work, using the ASS software designed especially for this purpose. This approach narrows the complexity of possible peptide ligands down to those occurring in nature. Phage displayed repertoires of proteins/domains obtained from different tissues and organisms are offered by various companies (e.g. Novagen, NEB). However, studies performed here revealed that large percentage of proteins/fragments displayed as fusions with the phage coat proteins, corresponds to non coding, nonsense DNA fragments. Such contaminations contribute to undesirable high phage background. Therefore the quality issue of the protein repertoire is of high relevance and preparation of custom libraries should be carefully controlled. The microarrays compatible with the process described here are not available commercially and must be fabricated from appropriate full length cDNA libraries. Such libraries are currently generated and fully characterized by the cDNA consortium of the German NGFN and comprehensive collections will be soon available. The final aspect of the entire high-throughput approach is the validation, segregation and storage of obtained data. A validation and statistical segregation process was developed in the context of this work, however more convenient databases and data mining tools must be found or developed in order to analyze and mine large amounts of data.

3.6 A novel structurally modified plastic surface for peptide synthesis and phage affinity selection

To apply the developed peptide-based phage panning in a systematic genome-wide ORF-scanning process would require the automation of as many steps as possible. High density arrays of peptide ligands attached to cellulose membranes have been successfully employed in this work for specific affinity selection of proteins displayed on a phage particle. However, utilization of peptides synthesized on such membranes requires separation (by cutting) of individual spots after the simultaneous multiple panning process and before the elution step. Otherwise, dissociated phages will diffuse to neighbouring spots and mix together. To avoid this, a novel structurally modified, non porous plastic surface (foil) was developed, which is suitable for spatially addressable peptide synthesis and phage elution (Patent DE 103 40 429.5: Frank *et al.*). Hydrophobic barriers, between hydrophilic acrylamide-patches, that were graft-polymerized on a planar polypropylene sheet, are designed to prevent contact of liquids containing phage populations dissociated from different peptides (Fig 3.6.1). This feature will also allow for an automation of the entire selection process using robotic pipetting tools of different array formats (i.e. 48, 96) and subsequent robotic transfer and propagation of the eluted phage populations in compatible 48 or 96 deep well culturing blocks.

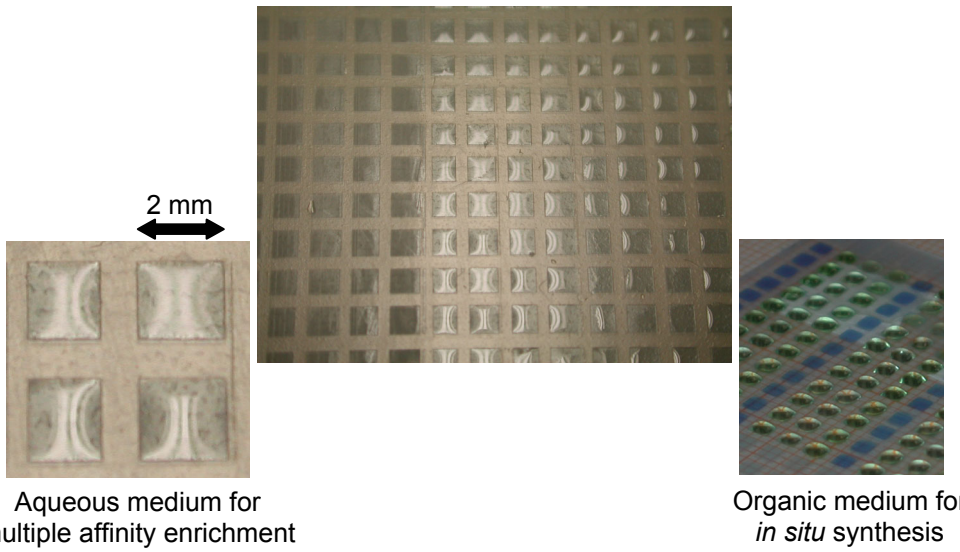
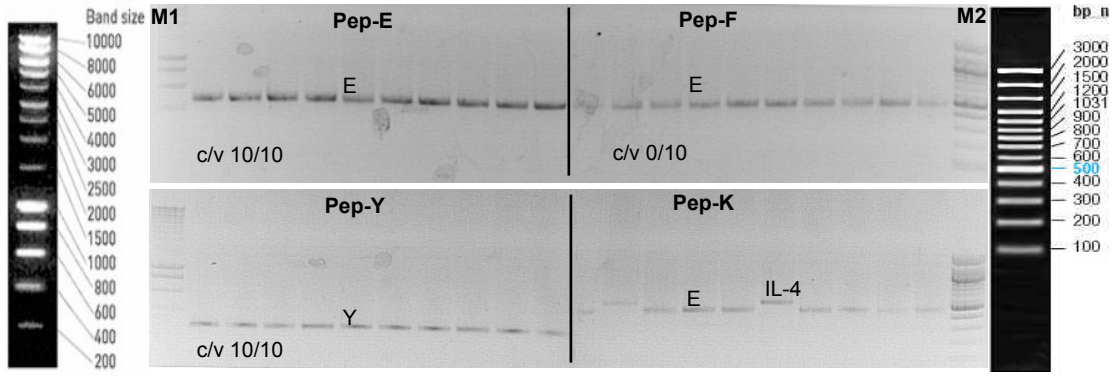


Fig 3.6.1: Pictures showing a novel plastic surface suitable for peptide synthesis and phage affinity selection experiments.

The main objective of this section was performance of various selection tests to study the binding abilities of phages displaying the three model protein domains (EVH1, WW/YAP, WW/FE65) to their corresponding peptides synthesized on this novel plastic surface. A series of patch arrays prepared with different chemistries all containing the model ligand peptide sequences (see Table 1.3.1 in section 1.3) were kindly provided by Dipl. Chem. Ing. Ulrike Beutling (personal communication GBF, Braunschweig). Initially, peptide arrays were subjected to phage affinity selection tests in order to choose the best performing one. Results described below were achieved using the UB194 array. For the preliminary tests, a solution containing high percentage of model phages was prepared by mixing them at a ratio of 1:100 with indifferent phages displaying mouse interleukin-4 (IL-4). Due to a high concentration of model phages within this mixture, large amounts of specifically bound phages were expected to be present within post-panning populations obtained from the peptide-patches. Before elution with guanidine solution, patches were cut out with the scalpel and placed in separate tubes, as described previously for the cellulose membranes (see 2.2.4.5). Following plating of the phages on a bacterial lawn, ten clones corresponding to each peptide, were randomly chosen for DNA isolation process followed by PCR amplification of DNA inserts and analysis of their sizes by AGE (Fig 3.6.2).

Array UB194: model phages diluted in IL-4 at ratio 1:100



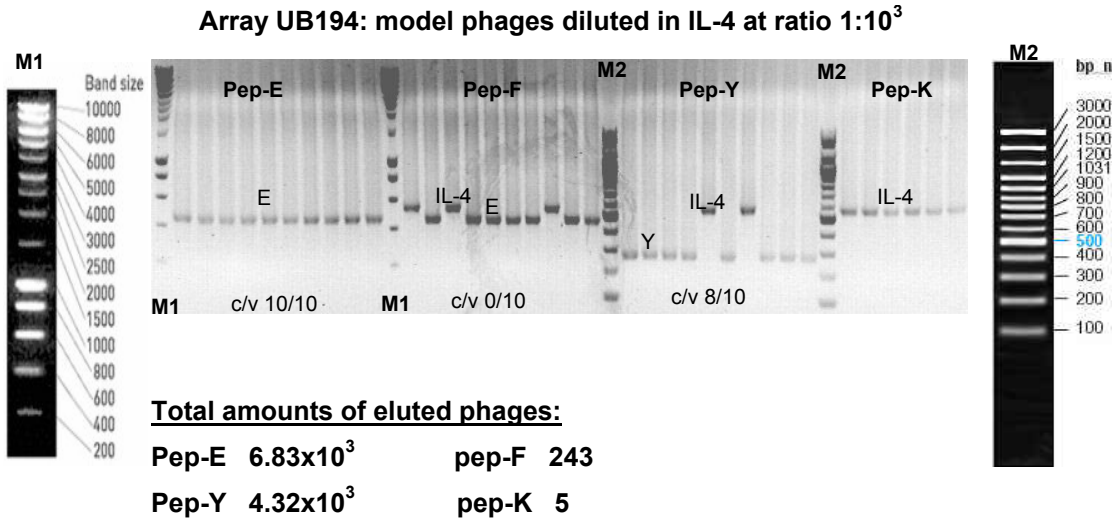
Total amounts of eluted phages

Pep-E	**2.5×10^4**	**pep-F**	**230**
Pep-Y	**6.1×10^3**	**pep-K**	**250**

Fig 3.6.2: Ethidium bromide stained 1 % AGE analysis of DNA inserts from selected phages that were enriched on model peptides synthesized on the patched plastic surface after panning with model phages diluted in IL-4 phage at ratio 1:100. Short name code: EVH1 (E), WW/FE65 (F); WW/YAP (Y); control (K); Interleukin-4 (IL-4); DNA markers: SMART (M1) (Eurogentec), XIII (M2) (Boehringer Mannheim); c/v correct/verified.

In general, much less phages were recovered from the patches as compared to the cellulose spots, which is expected as the resin material on the patches and the amount of immobilized peptides is also much less than the peptide/cellulose material of a spot (see Table 3.1.2 in section 3.1.2). Successful enrichment goes strict with a ten to one hundred fold increase in recovered phages: all phages obtained from the pep-E and pep-Y peptides contained the correct inserts. Quantity of phages eluted from the pep-F peptide resembled that from a control peptide (pep-K). Instead of the WW/FE65, only phages displaying the EVH1 domain were found within the population eluted from this peptide. Regarding the control peptide, phages displaying the IL-4 were expected to occur more frequently within the eluted population than the model phages, which were two folds of magnitude less represented within the pre-panning population. Nevertheless, phages displaying the EVH1 domain were also enriched on this patch. This supports the weak non-specific stickiness of the EVH1 phages, either to peptides or to the plastic support as observed in the previous chapter. Nevertheless the fact that the WW/FE65 domain was not enriched on its peptide, even despite of its high concentration within pre-panning mixture was surprising and may be a result of obscuring the pep-F ligands with the high excess of sticky EVH1 phages.

In order to test selection performances of the UB194 array when higher dilutions of model phages are applied, two additional pannings were carried out. New phage mixtures were prepared with the model phages diluted at ratios $1:10^3$ and $1:10^4$ in IL-4-phages. Panning and clone analysis were carried out as described above. As expected, higher dilutions of model phages resulted in slightly reduced amounts of phages eluted from their corresponding peptides, in both experiments. Phages recovered from pep-E, all contained DNA inserts characteristic for the EVH1 domain, in both experiments (Fig 3.6.3). Enrichment of phages displaying the WW/YAP domain on its peptide became less efficient, but still acceptable and is in agreement with previous results (see section 3.1). Eight and five correct inserts were found among ten analyzed clones for each experiment, respectively. Again it was observed that no WW/FE65 phage was enriched on its pep-F ligand. Analysis of phages recovered from panning with the $1:10^3$ mixture on the pep-F revealed seven clones containing inserts coding for the EVH1 domain and three clones containing inserts coding for IL-4. Only five phages in total, displaying only IL-4, were obtained from this peptide patch in the case of the $1:10^4$ mixtures. Obviously, the stickiness of the EVH1 phages is strictly concentration dependent and falls beyond the detection limit of the clone analysis method used. It is assumed that it will be still detectable with the more sensitive differential array method of the previous chapter. Only a few phages were eluted from the control peptides in both experiments when the sticky EVH1 phages are more diluted (Fig 3.6.3).



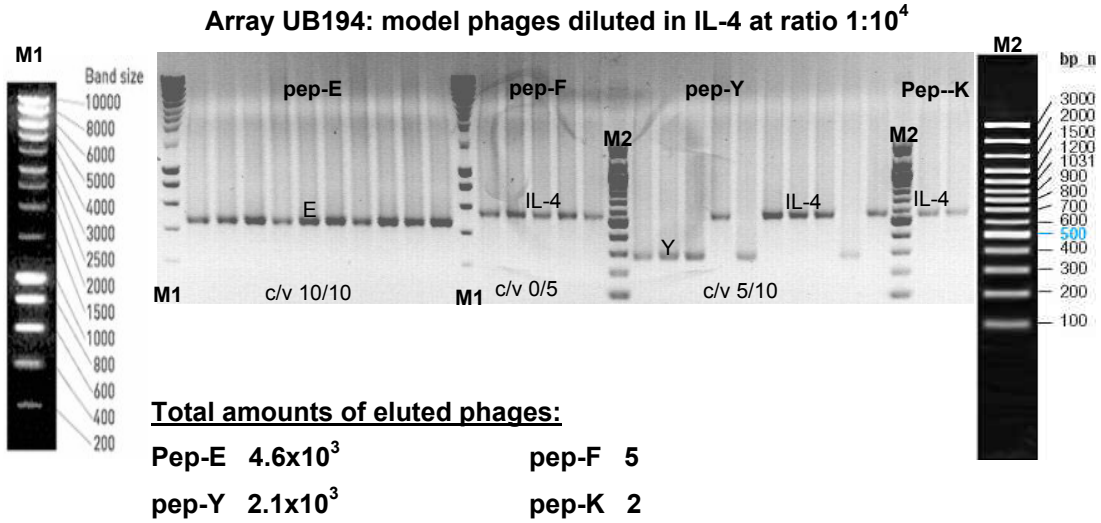


Fig 3.6.3: Ethidium bromide stained 1 % AGE analysis of DNA inserts from selected phages enriched on model peptides (10-15 aa) synthesized on patched plastic surface, after panning with the model phages diluted in IL-4 phage at ratios 1:10^3 and 1:10^4. Short name code: EVH1 (E), WW/FE65 (F); WW/YAP (Y); control (K); Interleukin-4 (IL-4); DNA markers: SMART (M1) (Eurogentec), XIII (M2) (Boehringer Mannheim); c/v correct/verified.

It could be argued that the observed failure in enrichment of the WW/FE65 could be due to either incorrect synthesis of its target peptide on the plastic surface or failure in display of functional protein domain on a phage particle caused by mutations within the phage chromosome. In order to check these possibilities, a similar control panning experiment utilizing the same model peptides synthesized on cellulose membrane was carried out as previously. Model phages were diluted again in IL-4-phage at ratio 1:10^3. As expected, amounts of peptide-spot eluted phages were higher than those observed for the patch arrays (Fig 3.6.4).

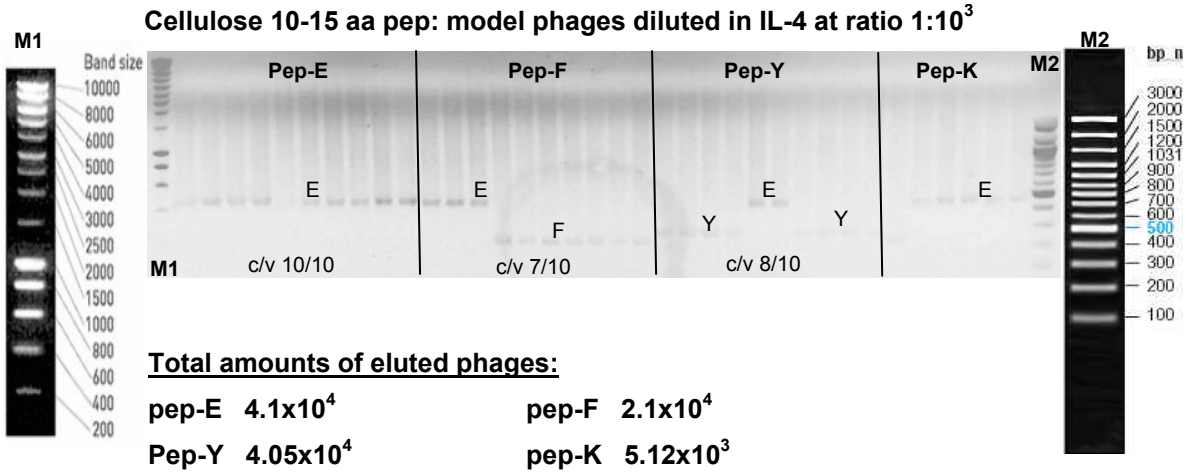


Fig 3.6.4: Ethidium bromide stained 1 % AGE analysis of DNA inserts from selected phages enriched on model peptides (10-15 aa) synthesized on cellulose membrane, after panning with model phages diluted in IL-4 at ratio 1:10^3. Short name code: EVH1 (E), WW/FE65 (F); WW/YAP (Y); control (K); Interleukin-4 (IL-4); DNA markers: SMART (M1) (Eurogentec), XIII (M2) (Boehringer Mannheim); c/v correct/verified.

Very efficient and specific enrichment of the EVH1 phages on their peptide-ligand (pep-E) was observed as before. Similarly, eight among ten phages achieved from pep-Y contained correct WW/YAP inserts. The remaining two phages displayed the EVH1 domain. Seven WW/FE65 phages were eluted from pep-F, although, three were presenting the EVH1 domain which clearly proves that the WW/FE65 domain is correctly displayed on a phage particle as a functional protein and rules out a failure of the patch enrichment because of non-presentation. The EVH1 phages were also found on the control spot, which confirms once more that these phages have some stickiness to other peptides or solid supports when occurring in higher concentrations in the panning libraries.

The performance of the patch array was compared with the respective experiments on cellulose spots by inspection of phage enrichment factors (see Table 3.6.1), calculated for the above bio-pannings, $1:10^3$ (patches and cellulose) and $1:10^4$ (patches) and the $1:10^4$ bio-panning described in section 3.1.2 (Table 3.1.3).

Except for the WW/FE65 phages, which were not successfully enriched on the patch arrays, a comparable enrichment was revealed for peptides pep-E and pep-Y in both cases when mixtures $1:10^3$ were applied for screening. In the case of bio-pannings with the mixture containing higher diluted model phages ($1:10^4$), performance of the patch array seemed even better, but more experiments need to be carried out in order to provide a reliable statistical comparison. The efficiency of phage enrichment on peptides synthesized on the patch support is very satisfactory. The advantage of this new type of non porous support is resulting reduction of total number of eluted phages with preservation of their specificity at the same time. Obviously, less phages that are irrelevant for binding to peptides are trapped on this solid support. This will be beneficial for future applications of such peptide arrays for large scale genome and proteome mapping processes.

Table 3.6.1: Comparison of enrichment factors calculated for phages selected on model peptides synthesized on acrylamide-patches and cellulose spots

Domain	Peptide	Acrylamide-patches UB194		Cellulose spots	
		Dilution ratio of model phages			
		$1:10^3$	$1:10^4$	$1:10^3$	$1:10^4$ (Table 3.1.3)
		enrichment factors			
EVH1	pep-E	1000	10000	1000	5250
WW FE65	pep-F	0	0	700	750
WW YAP	pep-Y	800	5000	800	4400

A further explanation for the failure of enriching WW/FE65 phages on the plastic patch bound pep-F could be a lower local density of the immobilized peptide compared to the peptide density on a cellulose spot leading to a reduced avidity in binding the multivalent T7 phages. Improvement of the process for patch generation, allowing a higher specific functionalization is in progress. Apart from the non-successful selection on the peptide-ligand pep-F synthesized on UB194 foil, the amounts of eluted phages, as well as selection specificities and enrichment factors observed for the two remaining peptides were very promising. A general lower amount of eluted phages, especially documented by the control and F peptides, suggested a significant reduction of unspecific background on this non-porous

surface. Nevertheless, more tests must be performed in order to prove a superior performance of this new type of peptide array.

Some important preliminary tests with a modified phage elution step using small volumes of 1 M guanidine solutions (2 – 5 μl) (Bialek, personal communication, GBF) added directly to the patches on the array, showed that these new peptide arrays are suitable for simultaneous parallel elution of phage populations from hydrophilic peptide patches separated via hydrophobic barriers. Future optimization of an elution protocol will allow for utilization of pipetting robots and high-throughput performance of the entire screening process. This promises an important role of the non-porous plastic surface in an automated panning process. Most of the other experimental steps in the screening process described in this thesis have been now implemented into high throughput screening formats using simultaneous parallel array methods and are carried out by only very few operations. These include the well established semi-automated peptide-SPOT synthesis method, as well as the massive production of cDNA biochips, both provide tools that are necessary for this approach and qualified to the repertoire of functional genomics and proteomics approaches. Similarly, the spot-wise parallel transfer of enriched phage populations from the peptide array sites to the DNA array analysis requires a series of standard operations, which need to be carried out by robotic instrumentation and evaluated by appropriate software in order to achieve an adequate throughput. Ultimately, a fully automated performance will definitely improve the quality and reliability of the whole process.

4. Conclusions and future prospects

This doctoral work aimed to establish a convenient methodology for the genome-wide mapping of interactions between protein domains and peptide ligands. The basic concept of a peptide array-based multiplexed affinity enrichment of protein domains displayed on bacteriophage capsids was successfully applied to three model domains and their respective known peptide ligands. Experimental comparisons of enrichment efficiencies of the model proteins using two different phage display systems, lambda and T7, clearly favored the T7 phage as more suitable for purposes of this project. Several different approaches were stepwise evaluated during this work, in order to find a suitable methodology for identification of phages specifically enriched on target peptides. These steps include the classical analysis of phage clones based on PCR, AGE and DNA sequencing, as well as the more sophisticated template assisted iterative screening (TAIS) methodology. None of these approaches, however, provided a satisfactory performance when used in the context of a highly diverse phage display library screened versus an array of peptides immobilized on a cellulose membrane. Finally, a combination of three well characterized methodologies led to an effective process, entirely based on high-throughput biochip technologies. In the first step of this process, a phage display library of protein domains will be “panned” over repertoires of synthetic peptides immobilized as arrays on a solid support. Due to the advantageous feature of phage display technology, which is a physical linkage of phenotype with genotype, displayed proteins/domains will be identified via differential hybridization of their labeled pre- and post-panning DNA inserts with an array of complementary DNA-probe sequences printed on a modified glass surface. Differences between the amounts of DNA inserts within both populations will be reflected by different values of fluorescent signal intensities, achieved from DNA spots after the hybridization process. Overrepresentation of certain open reading frame (ORF) DNA inserts within the post-panning population with respect to the pre-panning population will indicate enrichment of the expressed protein presented on the phage capsid. This process was successfully implemented for the three model domains. A special mode of microarray data analysis, based on restrictive thresholds and statistical biochip comparisons, was developed to distinguish between specifically and non-specifically bound candidates. Using this approach, screening of the entire human brain phage display cDNA library, succeeded in selection and identification of WW domain containing proteins (AIP4, AIP5), interacting specifically with proline-rich target peptides covalently attached to the cellulose membrane. Although information obtained from such screenings still needs to be validated for real biological evidence by using independent methodologies, the developed process was shown to narrow a large population of phage displayed proteins down to a sub-set containing meaningful potential binding partners.

The developed process can be applied to diverse types of libraries, prepared from DNA or cDNA preparations of tissues or organisms of healthy or pathogenic origin, to be screened against any type of dedicated or non-dedicated peptide repertoires. Identification of novel interactions will support the elucidation of a potential biological function of both interacting proteins and could help to address these to certain signaling pathways or biochemical processes. Modern bioinformatics tools based on database searching and analysis of protein networks greatly enhanced the profiling of potential relations between identified partners and other proteins that are part of interaction networks connected to a certain biological process.

To achieve the anticipated genome-wide mapping of peptide-mediated protein-protein interactions, an advanced high-throughput performance of the process is highly desirable. One of the requirements of such a process is an optimized design of peptide arrays. Generic peptide libraries as well as a genome-wide overlapping peptide scan covering all (e.g. human) ORFs would be adequate peptide collections for this approach, but these could be too large for paractical work. On the other hand, utilizing combinatorial search strategies, less voluminous libraries of partially randomized peptide pools (10^3-10^6 components) with built-in information about recognition motives could be applied to reduce the number of peptide preparations. In this work, a significant reduction of the number of peptides for an ORF scanning approach was achieved by application of an Accessible Surface Scanner software tool (ASS). The underlying algorithm was developed for prediction of surface exposed continuous, peptidic segments which are considered to posess the binding sites responsible for interaction with other proteins. Sequences predicted as buried within the core of a protein were assumed to be irrelevant for binding to other protein domains and therefore eliminated. Analyses of several protein sequence containing known, linear recognition motifs responsible for binding to various protein modules revealed that in the majority of cases, these motifs were correctly predicted by the ASS software tool as surface exposed. Subsequently, *in vitro* phage selection experiments performed on peptides selected with the help this software, proved its acceptable performance. Moreover, ASS allows for rapid simultaneous processing of thousands of protein sequences retrieved from data bases. Compatibility of peptide output data with the software that is controlling the robotic peptide synthesis process is another advantageous feature supporting the high performance of the entire process.

A second requirement for an advanced genome spanning protein interaction screening process is its automation. Many steps that are important for preparation of materials necessary for screenings are already performed by robotic systems, including SPOT peptide synthesis and fabrication of DNA microarray biochips. Thus, the automation of the phage affinity selection process would be highly desirable. A first step toward this goal was the successful evaluation of a novel structurally modified, non porous plastic surface, which is

suitable for a spatially addressable peptide synthesis and simultaneous multiple phage panning. The structure of this array provides peptides synthesized on hydrophilic patches surrounded by hydrophobic barriers. This allows for automatic post-panning elution of the enriched phage populations, excluding contact and mutual contamination of phage containing liquids from neighbouring peptide patches. Automation of panning and elution steps would result in a rapid increase of process performance and generation of large amounts of data. Model peptides synthesized on this new support were successfully used for specific phage enrichment. These studies revealed significant advantages of the new peptide array over the classical cellulose support for peptide arrays synthesis. Proposals for further improvements of this new type of peptide array have been worked out. Finally, design and development of suitable databases for data storage, as well as sophisticated bioinformatics tools for data mining and comparison are next in the line for the future work and have to be advanced in order to reach a fully automated and functional high-throughput methodology for peptide-based proteome mapping.

5. Literature

Asland, R., Abrams, C., Ampe, C., Ball, L.J., Bedford., M.T, Cesareni, G., Gimona, M., Hurley, J.H., Jarchau, T., Lehto, V.P., Lemmon, M.A., Linding, R., Mayer, B.J, Nagai, M., Sudol, M., Walter, U., Winder, S.J. Normalization of nomenclature for peptide motifs as ligands of modular protein domains. *FEBS Lett*. 2002, 513, 1:141-4

Adey, N.B., Sparks, A.B., Beasley, J. and Kay, B.K. Construction of random peptide libraries in bacteriophage M13. In: Phage display of peptides and proteins: A laboratory manual (Kay, B.K., Winter, J. and McCafferty, J., eds) *Academic Press, San Diego*, 1996, 67-78

Alexandropoulos, K., Cheng, G. and Baltimore, D. Proline-rich sequences that bind to Src homology 3 domains with individual specificities. *Proc Natl Acad Sci USA*, 1995, 92: 3110-3114

Allen, J. B., Walberg, M.W., Edwards, M.C. and Elledge, S.J. Finding prospective partners in the library: the two-hybrid system and phage display find a match. *Trends Biochem. Sci.* 1995, 20: 511–516

Atassi, M.Z. The antigenic structure of myoglobin and initial consequences of its precise determination. *CRC Crit Rev Biochem*, 1979, 6, 4: 337-369

Atassi, M.Z and Kurisaki, J. A novel approach for localization of the continuous protein antigenic sites by comprehensive synthetic surface scanning: antibody and T-cell activity to several influenza hemagglutinin synthetic sites. *Immunol Commun*, 1984, 13, 6: 539-551

Ayyadevara, S., Thaden, T.J. and Shmookler Reis R.J. Anchor polymerase chain reaction display: a high-throughput method to resolve, score, and isolate dimorphic genetic markers based on interspersed repetitive DNA elements. *Anal. Biochem*, 2000, 284, 19-28

Ball, L.J., Jarchau, T., Oschkinat, H. and Walter, U. EVH 1 domains: structure, function and interactions. *Febs Letters*, 2002, 513, 1: 45-52

Ball, L.J., Kühne, R., Hoffmann, B., Häfner, A., Schmieder, P., Volkmer-Engert, R., Hof, M., Wahl, M., et al. Dual epitope recognition by the VASP EVH1 domain modulates polyproline ligand specificity and binding affinity. *EMBO J*, 2000, 19: 4903-4914

Barfod, E.T., Zheng, Y., Kuang, W.J., Hart, M.J., Evans, T., Cerione, R.A. and Ashkenazi, A. Cloning and expression of a human CDC42 GTPase-activating protein reveals a functional SH3-binding domain. *J Biol Chem*, 1993, 268, 35: 26059-26062

Barnett, P., Bottger, G., Klein, A.T.J., Tabak, H.F. and Distel, B. The peroxisomal membrane protein Pex13p shows a novel mode of SH3 interaction. *EMBO J,* 2000, 19: 6382-6391

Barnes, W.M. PCR amplification of up to 35-kb DNA with high fidelity and high yield from A bacteriophage templates. *Genetics*, 1994, 91: 2216-2220

Barzik, M., Carl, U.D., Schubert, W.-D., Frank, R., Wehland, J. and Heinz, D.W. The N-terminal domain of Homer/Vesl is a new class II EVH1 domain. *J Mol Biol,* 2001, 309: 155-169

Bedford, M.T., Sarbassova, D., Xu, J., Leder, P. and Yaffe, M.B. A novel Pro-Arg motif recognized by WW domains. *J Biol Chem,* 2000, 275: 10359-10369

Beebe, K.D., Wang, P., Arabaci, G. and Pei, D. Determination of the binding specificity of the SH2 domains of protein tyrosine phosphatase SHP-1 through the screening of a combinatorial phosphotyrosyl peptide library. *Biochemistry*, 2000, 39: 13251-13260

Berry, D.M., Nash, P., Liu, S. K.-W., Pawson, T. and McGlade, C.J. A high-affinity Arg-X-X-Lys SH3 binding motif confers specificity for the interaction between Gads and SLP-76 in T cell signaling. *Curr Biol*, 2002, 12: 1336-1341

Bialek, K., Swistowski, A.. and Frank, R. Epitope-targeted proteome analysis: towards a large-scale automated protein-protein-interaction mapping utilizing synthetic peptide arrays. *Anal Bioanal Chem,* 2003, 376: 1006-1013

Bialkowska, A. and Kurlandzka, A. Proteins interacting with Lin 1p, a putative link between chromosome segregation, mRNA splicing and DNA replication in *Saccharomyces cerevisiae*. *Yeast*, 2002, 19, 15: 1323-1333

Birnoim, H.C. and Doly, J. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucl. Acids Res*, 1979, 7: 1513-1522

Borinstein, S.C., Hyatt, M.A., Sykes, V.W., Straub, R.E., Lipkowitz, S., Boulter, J. and Bogler, O. SETA is a multifunctional adapter protein with three SH3 domains that binds Grb2, Cbl, and the novel SB1 proteins. *Cell Signal*, 2000, 12: 769-779

Bowtel and Sambrook, J. DNA Microarrays: A Molecular Cloning Manual, Cold Spring Harbor Laboratory Press, 2002

Carl, U. D., Pollmann, M., Orr, E., Gertler, F.B., Chakraborty, T. and Wehland, J. Aromatic and basic residues within the EVH1 domain of VASP specify its interaction with proline-rich igands. *Current Biology*, 1999, 9, 13: 715-718

Castagnoli, L., Zuccconi, A., Quondam, M., Rossi, M., Vaccaro, P., Panni, S., Paoluzi, S., Santonico, E., Dente, L. and Cesareni, G. Alternative Bacteriophage Display Systems *Comb. Chem. High Throughput Screen*, 2001, 4: 121–133

Cestra, G., Castagnoli, L., Dente, L., Minenkova, O., Petrelli, A., Migone, N., Hoffmuller, U., Schneider-Mergener, J., et al. The SH3 domains of endophilin and amphiphysin bind to the proline- rich region of synaptojanin 1 at distinct sites that display an unconventional binding specificity. *J Biol Chem*, 1999, 274: 32001-32007

Chan, D.C., Bedford, M.T. and Leder, P. Formin binding proteins bear WWP/WW domains that bind proline-rich peptides and functionally resemble SH3 domains. *EMBO J*, 1996, 15: 1045-1054

Chan, W.C. and White, P.D. Fmoc solid phase peptide synthesis: a practical approach. *Oxford University Press, New York,* 2000

Chang, B.Y., Conroy, K.B., Machleder, E.M. and Cartwright, C.A. RACK1, a receptor for activated C kinase and a homolog of the β subunit of G proteins, inhibits activity of src tyrosine kinases and growth of NIH 3T3 cells. *Mol Cell Biol,* 1998, 18: 3245-3256

Chatellard-Causse, C., Blot, B., Cristina, N., Torch, S., Missotten, M. and Sadoul, R. Alix (ALG-2-interacting protein X), a protein involved in apoptosis, binds to endophilins and induces cytoplasmic vacuolization. *J Biol Chem,* 2002, 277: 29108–29115

Cheadle, C., Ivashchenko, Y., South, V., Searfoss, G.H., French, S., Howk, R., Ricca, G.A. and Jaye, M. Identification of a Src SH3 domain binding motif by screening a random phage display library. *J Biol Chem,* 1994, 269: 24034-24039

Chen, B-Y and Janes, H. W. PCR Cloning Protocols. Methods in Molecular Biology*, 2002, 192*

Chen, E.H., Fre, S., Slepnev, V.I., Capua, M.R., Takei, K., Butler, M.H., Di Fiore, P.P. and De Camilli, P. psin is an EH-domain-binding protein implicated in clathrin-mediated endocytosis. *Nature*, 1998, 394, 6695: 793-797

Chen, H.I., Einbond, A., Kwak, S.-J., Linn, H., Koepf, E., Peterson, S., Kelly, J.W. and Sudol, M. Characterization of the WW domain oh human yes-associated protein and its polyproline-containing ligands. *J Boil Chem*, 1997, 272: 17070-17077

Chen, H.I. and Sudol, M. The WW domain of Yes-associated protein binds a proline-rich ligand that differs from the consensus established for Src homology 3-binding modules. *Proc Natl Acad Sci USA*, 1995, 92: 7819-7823

Cheng, J., Moyer, B.D., Milewski, M., Loffing, J., Ikeda, M., Mickle, J.E., Cutting, G.R., Li, M., et al. A Golgi-associated PDZ domain protein modulates cystic fibrosis transmembrane regulator plasma membrane expression. *J Biol Chem*, 2002, 277: 3520–3529

Chou, P.Y. and Fasman G.D. Prediction of the secondary structure of proteins from their amino acid sequence. *Advances in Enzymology and Related Areas of Molecular Biology,* 1978, 47: 45–148

Cicchetti, P., Mayer, B.J., Thiel, G. and Baltimore, D. Identification of a protein that binds to the SH3 region of Abl and is similar to Bcr and GAP-rho. *Science*, 1992, 257: 803-806

Cline, J., Braman, J.C. and Hogrefe, H.H. PCR fidelity of pfu *DNA polymerase and other thermostable DNA polymerases.* Nucleic Acids Res*, 1996, 24,18: 3546-3551*

Clackson, T., Hoogenboom, H.R., Griffiths, A.D. and Winter, G. Making antibody fragments using phage display libraries. *Nature*, 1991, 352: 624–628

Cochrane, D., Webster, C., Masih, G. and McCafferty, J. Identification of natural ligands for SH2 domains from a phage display cDNA library. *J Mol Biol*, 2000, 297: 89–97

Dente, L., Vetriani, C., Zucconi, A., Pelicci, G., Lanfrancone, L., Pelicci, P.G. and Cesareni, G. Modified phage peptide libraries as a tool to study specificity of phosphorylation and recognition of tyrosine containing peptides. *J Mol Biol*, 1997, 269: 694-703

Diehl, F., Beckmann, B., Kellner, M. Hauser, N.C., Diehl, S. and Hoheisel, J. D. Manufacturing DNA microarrays from unpurified PCR products. *Nucleic Acids Res*, 2002, 30, 16: e79

Diehl, F., Grahlmann, S., Beier, M. and Hoheisel, J. D. Manufacturing DNA microarrays of high spot homogeneity and reduced background signal. *Nucleic Acids Res*, 2001, 29, 7: 38

Di Vignano, A.T., Di Zenzo, G., Sudol, M., Cesareni, G. and Dente, L. Contribution of the different modules in the utrophin carboxy-terminal region to the formation and regulation of the DAP complex. *FEBS Lett,* 2000, 471: 229-234

Dower, W.J., Miller, J.F. and Ragsdale, C.W. High efficiency transformation of *E. coli* by high voltage electroporation. *Nucleic Acids Res*, 1988, 16: 6127-6145

Drees, B., Friederich, E., Fradelizi, J., Louvard, D., Beckerle, M.C. and Golsteyn, R.M. Characterization of the interaction between zyxin and members of the Ena/vasodilator-stimulated phosphoprotein family of proteins. *J Biol Chem,* 2000, 275: 22503-22511

Durocher, D., Taylor, I.A., Sarbassova, D., Haire, L.F., Westcott, S.L., Jackson, S.P., Smerdon, S.J. and Yaffe, M.B. The molecular basis of FHA domain: phosphopeptide binding specificity and implications for phospho-dependent signaling mechanisms. *Mol Cell,* 2000, 6: 1169-1182

Emini, E.A., Hughes, J.V., Perlow, D.S. and Boger J. Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. *Journal of Virology*, 1985, 55: 836–839

Ermekova, K.S., Zambrano, N., Linn, H., Minopoli, G., Gertler, F., Russo, T. and Sudol, M. The WW domain of neural protein FE65 interacts with proline-rich motifs in Mena, the mammalian homolog of *Drosophila* enabled. *J Biol Chem,* 1997, 272: 32869-32877

Espanel, X. and Sudol, M. A. single point mutation in a group I WW domain shifts its specificity to that of group II WW domains. *J Biol Chem*, 1999, 274: 17284-17289

Espanel, X. and Sudol, M. Yes-associated protein and p53-binding protein-2 interact through their WW and SH3 domains. *J Biol Chem*, 2001, 276: 14514-14523

Espejo, A., Côté, J., Bednarek, A., Richard, S. and Bedford, M.T. A protein-domain microarray identifies novel protein-protein interactions. *Biochem J*, 2002, 367: 697-702

Fare, T.L, Coffey, E.M, Dai, H., He, Y.D., Kessler, D.A., Kilian, K.A., Koch, J.E., LeProust, E., Marton, M.J., Meyer, M.R., Stroughton, R.B., Tokiwa, G.Y. and Wang, Y. Effects of atmospheric ozone on microarray data quality. *Anal Chem*, 2003, 75, 17: 4672-4675

Fazi, B., et al. Unusual binding properties of the SH3 domain of the yeast actinbinding protein Abp1: structural and functional analysis. *J. Biol. Chem.* 2002, 277: 5290–5298

Felici, F., Castagnoli, L., Musacchio, A., Jappelli, R. and Cesareni, G. Selection of antibody ligands from a large library of oligopeptides expressed on a multivalent exposition vector. *J Mol Biol*, 1991, 222: 301-310

Fisch, I. Peptide display in functional genomics. *Comb. Chem. High Throughput Screen*, 2001, 4: 157–169

Flynn, D.C., Leu, T.H., Reynolds, A.B. and Parsons, J.T. Identification and sequence analysis of cDNAs encoding a 110-kilodalton actin filament-associated pp60src substrate. *Mol Cell Biol.*, 1993, 13, 12: 7892–7900

Frackman, S., Kobs, G., Simpson, D. and Storts,D. Betaine and DMSO. Enhancing agents for PCR. Promega Notes, 1998, 65: 27

Frank, R. Spot-synthesis: an easy technique for positionally addressable, parallel chemical synthesis on a membrane support. *Tetrahedron*. 1992, 48, 42: 9217-9232

Frank, R. High-density synthetic peptide microarrays: emerging tools for functional genomics and proteomics. *Comb Chem High Throughput Screen*, 2002, 5: 429-440

Frank, R. The SPOT-synthesis technique. Synthetic peptide arrays on membrane supports - principles and applications. *J Immunol Mehodst*, 2002, 267: 13-26

Frank *et.al*. Patent DE 103 40 429.5: Hydrophober Gegenstand mit Raster hydrophiler Bereiche, dessen Herstellung und Verwendung. 2003

Frank, R. and Overwin, H. SPOT-synthesis. Epitope analysis with arrays of synthetic peptides prepared on cellulose membranes. In: Morris, G.E. (Ed.), Methods in Molecular Biology. Epitope Mapping Protocols. *Humana Press, Totowa,* 1996, 66: 149-169

Fuchs, A., Dagher, M.-C. and Vignais, P.V. Mapping the domains of interaction of $p40^{phox}$ with both $p47^{phox}$ and $p67^{phox}$ of the neutrophil oxidase complex using the two-hybride system. *J Biol Chem*, 1995, 270: 5695-5697

Fuh, G., Pisabarro, M.T., Li, Y., Quan, C., Lasky, L.A. and Sidhu, S.S. Analysis of PDZ domain-ligand interactions using carboxyl-terminal phage display. *J Biol Chem*, 2000, 275: 21486-21491

Garcia, R.A.G, Vasudevan, K. and Buonanno, A. The neuregulin receptor ErbB-4 interacts with PDZ-containing proteins at neuronal synapses. *PNAS*, 2000, 97, 7: 3596-3601

Gary, R. and Bretscher A. Ezrin self-association involves binding of an N-terminal domain to a normally masked C-terminal domain that includes the F-actin binding site. *Mol Biol Cell*, 1995, 6, 8: 1061-1075

Gavin, A.C., Bosche, M., Krause, R., Grandi, P., Marzioch, M., et al. Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature, 2002, 415: 141–147

Gee, S.H., Sekely, S.A., Lombardo, C., Kurakin, A., Froehner, S.C. and Kay, B.K. Cyclic peptides as non-carboxyl-terminal ligands of syntrophin PDZ domains. *J Biol Chem*, 1998, 273: 21980-21987

Giot, L., et al. A protein interaction map of *Drosophila melanogaster*. *Science*, 2003, 302, 1727–1736

Giovanni Cesareni (Editor), Mario Gimona (Editor), Marius Sudol (Editor), Michael Yaffe (Editor) Modular Protein Domains ISBN: 3-527-30813-X December 2004

Gohlke, H., Kuhn, L.A. and Case, D.A. Change in protein flexibility upon complex formation: analysis of Ras-Raf using molecular dynamics and a molecular framework approach. *Proteins*, 2004, 56, 2: 322-337

Gout, I., Dhand, R., Hiles, I.D., Fry, M.J., Panayotou, G., Das, P., Truong, O., Totty, N.F., Hsuan, J., Booker, G.W., et al.The GTPase dynamin binds to and is activated by a subset of SH3, domains. *Cell*, 1993, 75, 1: 25-36

Grabs, D., Slepnev, V.I., Sogyang, Z., David, C., Lynch, M., Cantley, L.C. and De Camilli, P. The SH3 domain of amphiphysin binds the proline-rich domain of dynamin at a single site that defines a new SH3 binding consensus sequence. *J Biol Chem*, 1997, 272: 13419-13425

Gram, H., Schmitz, R., Zuber, J.F. and Baumann, G. Identification of phosphopeptide ligands for the Src-homology 2 (SH2) domain of Grb2 by phage display. *Eur J Biochem*, 1997, 246, 633-637

Haffner, C., Takei, K., Chen, H., Ringstad, N., Hudson, A., Butler, M.H., Salcini, A.E., Di Fiore, P.P. and De Camilli, P. Synaptojanin 1: localization on coated endocytic intermediates in nerve terminals and interaction of its 170 kDa isoform with Eps15. *FEBS Lett*, 1997, 419, 2-3: 175-180

Hart, C.P., Martin, J.E., Reed, M.A., Keval, A.A., Pustelnik, M.J., Northrop, J.P., Patel, D.V. and Grove, J.R. Potent inhibitory ligands of the GRB2 SH2 domain from recombinant peptide libraries. *Cell Signal*, 1999, 11: 453-464

Hiipakka, M., Poikonen, K. and Saksela, K. SH3 domains with high affinity and engineered ligand specificities targeted to HIV-1 Nef. *J Mol Biol*, 1999, 293: 1097-1106

Hoffmüller, U., Russwurm, M., Kleinjung, F., Ashurt, J., Oschkinat, H., Volkmer-Engert, R., Koesling, D. and Schneider-Mergener, J. Interaction of a PDZ protein domain with a synthetic library of all human protein C termini. *Angew Chem Int Ed*, 1999, 38: 2000-2004

Ho, Y., Gruhler, A., Heilbut, A., Bader, G.D., Moore, L. et al. (2002) Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry. *Nature* 415: 180–183

Hoogenboom, H.R., de Bruine A.P., Hufton, S.E., Hoet, R.M., Arends, J-W. and Roovers, R.C. Antibody phage display and its applications. 1998. *Immunotechnology.* 4: 1–20

Hu, H., Columbus, J., Zhang, Y., Wu, D., Lian, L., Yang, S., Goodwin, J., Luczak, C., et al. A map of WW domain family interactions. *Proteomics*, 2004, 4: 1-13

Huang, M.M., Arnheim, N. and Goodman M.F. Extension of base mispairs by Taq DNA polymerase: implications for single nucleotide discrimination in PCR. *Nucleic Acids Res*, 1992, 20, 17: 4567-4573

Hultschig, C. Two dimensional screening: towards establishing a novel technique to study biomolecular interactions. *Phd. thesis,*
http://www.biblio.tubs.de/ediss/data/20000207a/20000207a.html 2000

Howell, B.W., Lanier, L.M., Frank, R., Gertler, F.B. and Cooper, J.A. The disabled 1 phosphotyrosine-binding domain binds to the internalisation signals of transmembrane glycoproteins and to phospholipids. *Mol Cell Biol*, 1999, 19, 5179-5188

Igarashi, K., Shigeta, K., Isohara, T., Yamano, T. and Uno, I. Sck interacts with KDR and Flt-1 via its SH2 domain. *Biochem Biophys Res Comm* 1998, 251, 77-82

Ikeda, M., Ikeda, A., Longan, L. C. and Longnecker, R. The Epstein–Barr virus latent membrane protein 2A PY motif recruits WW domain-containing ubiquitin–protein ligases. *Virology* 2000, 268, 178–191

Inouye, H. and Kirschner, D.A. Folding and function of the myelin proteins from primary sequence data. *J Neurosci Res*, 1991, 28, 1, 1-17

Ito, T., Chiba, T., Ozawa, R., Yoshida, M., Hattori, M. and Sakaki, Y. A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc. Natl. Acad. Sci. U S A* 2001, 98, 4569–4574

Jameson, B.A. and Wolf, H. The antigenic index: a novel algorithm for predicting antigenic determinants. *Comput Appl Biosci* 1988, 4, 181–186

Janin, J., Wodak, S,. Levitt, M. and Maigret. B. Conformation of amino acid side-chains in proteins. 1978. *J Mol Biol*, 125: 357-386

Johnsson, N. and Varshavsky, A. Split ubiquitin as a sensor of protein interactions in vivo. *Proc. Natl. Acad. Sci. USA* 1994, 91, 10340–10344

Kapeller, R., Prasad, K.V., Janssen, O., Hou, W., Schaffhausen, B.S., Rudd, C.E. and Cantley, L.C. Identification of two SH3-binding motifs in the regulatory subunit of phosphatidylinositol 3-kinase. *J Biol Chem*, 1994, 269, 3: 1927-1933

Karlsson, T., Songyang, Z., Landgren, E., Lavergne, C., Di Fiore, P.P., Anafi, M., Pawson, T., Cantley, L.C., Claesson-Welsh, L. and Welsh, M. Molecular interactions of the Src homology 2 domain protein Shb with phosphotyrosine residues, tyrosine kinase receptors and Src homology 3 domain proteins. *Oncogene*, 1995, 10, 8, 1475-1483

Karplus, P. and Shulz, G. Prediction of chain flexibility in proteins: a tool for the selection of peptide antigens. *Naturwissenschaften*, 1985, 72: 212–213

Kasanov, J., Pirozzi, G., Uveges, A.J., and Kay, B.K. Characterizing Class I WW domains defines key specificity determinants and generates mutant domains with novel specificities. *Chem Biol* 2001, 8: 231-241

Kay, B.K., Williamson, M.P. and Sudol, M. 41 The importance of being proline: the interaction of proline-rich motifs in signaling proteins with their cognate domains. *FASEB J*, 2000, 14: 231- 232

Kelly, M.A., Liang, H., Sytwu, I.-I., Vlattas, I., Lyons, N.L., Bowen, B.R. and Wennogle, L.P. Characterization of SH2-ligand interactions via library affinity selection with mass spectrometric detection. *Biochemistry*, 1996, 35: 11747-11755

Kessels, H.W.H.G., Ward, A.C. and Schumacher, T.N.M. Specificity and affinity motifs for Grb2 SH2-ligand interactions. *Proc Natl Acad Sci USA*, 2002, 99: 8524-8529

Kikonyogo, A., Bouamr, F., Vana, M.L., Xiang, Y., Aiyar, A., Carter, C. and Leis, J. Proteins related to the Nedd4 family of ubiquitin protein ligases interact with the L domain of Rous sarcoma virus and are required for gag budding from cells. *Proc Natl Acad Sci USA*, 2001, 98: 11199–11204

Kirsch, K.H., Georgescu, M.-M., Ishimaru, S. and Hanafusa, H. CMS: an adapter molecule involved in cytoskeletal rearrangements. *Proc Natl Acad Sci USA*, 1999, 96: 6211-6216

Knudsen, B.S., Feller, S.M. and Hanafusa, H. Four proline-rich sequences of the guanine-nucleotide exchange factor C3G bind with unique specificity to the first Src homology 3 domain of Crk. *J Biol Chem*, 1994, 269, 52: 32781-32787

Kofler, M., Heuer, K., Zech, T. and Freund, C. Recognition sequences for the GYF domain reveal a possible spliceosomal function of CD2BP2. *J Biol Chem*, 2004, 279, 27: 28292-28297

Kramer, A., Reineke, U., Dong, L., Hoffmann, B., Hoffmüller, U., Winkler, D., Volkmer-Engert, R. and Schneider-Mergener, J. SPOT synthesis: observations and optimizations. *J Pept Res*, 1999, 54: 319-327

Kurakin, A. and Bredesen, D. Target assisted iterative screening reveals novel interactors for PSD95, Nedd4, Src, Abl and Crk proteins. *J. Biomol. Struct. Dyn.*, 2002, 19: 1015–1029

Kurakin, A.V., Wu, S. and Bredesen, D.E. Atypical recognition consensus of CIN85/SETA/Ruk SH3 domains revealed by target-assisted iterative screening. *J Biol Chem*, 2003, 278: 34102-34109

Kyte, J. and Doolittle, R. F. A simple method for displaying the hydropathic character of a protein. *J Mol Biol*, 1982, 157: 105–132

Lam, K.S., Liu, R., Miyamoto, S., Lehman, A.L. and Tuscano, J.M. Applications of one-bead one-compound combinatorial libraries and chemical microarrays in signal transduction research. *Acc Chem Res*, 2003, 36: 370-377

Landgraf, C., Panni, S., Montecchi-Palazzi, L., Castagnoli, L., Schneider-Mergener, J., Volkmer-Engert, R. and Cesareni, G. Protein interaction networks by proteome peptide scanning. *PLoS Biol*, 2004, 2: 94-103

Laura, R.P, Witt, A.S., Held, H.A., Gerstner, R., Deshayes, K., Koehler, M.F.T., Kenneth S. Kosik, K.,S., Sidhu, S.S. and Lasky, L.A. The Erbin PDZ domain binds with high affinity and specificity to the carboxyl termini of δ-catenin and ARVCF. *Biol Chem*, 2002, 277, 15: 12906-1291

Li, S.-C., Songyang, Z., Vincent, S.J.F., Zwahlen, C., Wiley, S., Cantley, L., Kay, L.E., Forman-Kay, J., et al. High-affinity binding of the *Drosophila* Numb phosphothyrosine-binding domain to peptides containing a Gly-Pro-(p)Tyr motif. *Proc Natl Acad Sci USA*, 1997, 94: 7204–7209

Liao, H., Yuan, C., Su, M.-I., Yongkiettrakul, S., Qin, D., Li, H., Byeon, I.-J.L., Pei, D., et al. Structure of the FHA1 domain of yeast Rad53 and identification of binding sites for both FHA1 and its target protein Rad9. *J Mol Biol*, 2000, 304: 941-951

Linn, H., Ermekova, K.S., Rentschler, S., Sparks, A.B., Kay, B.K. and Sudol, M. Using molecular repertoires to identify high-affinity peptide ligands of the WW domain of human and mouse YAP. *Biol Chem*, 1997, 378: 531-537

Lipshutz, R.J., Fodor, S.P., Gingeras, T.R. and Lockhart, D.J. High density synthetic oligonucleotide arrays. *Nat Genet*, 1999, 21, 1 Suppl: 20-24

Lipshutz, R.J, Morris, D., Chee, M., Hubbell, E., Kozal, M.J., Shah, N., Shen, N., Yang, R., Fodor, S.P. Using oligonucleotide probe arrays to access genetic diversity. *Biotechniques*, 1995, 19, 3: 442-447

Liu, R., Enstrom, A.M. and Lam, K.S. Combinatorial peptide library methods for immunobiology research. *Exp Hematol*, 2003, 31: 11-30

Liu, S. K., Fang, N., Koretzky, G.A., McGlade, C.J. The hematopoietic specific adaptor protein gads functions in T-cell signaling via interactions with the SLP-76 and LAT adaptors. *Curr. Biol.,* 1999, 9: 67–75

Liu, S.K. and McGlade, C.J. Gads is a novel SH2 and SH3 domain-containing adaptor protein that binds to tyrosine-phosphorylated Shc. *Oncogene*, 1998, 17: 3073-3082

Lucke, S., Grunwald, T., Uberla, K. Reduced mobilization of Rev-responsive element-deficient lentiviral vectors. *Virol*., 2005, 79, 14:9359-62

MacBeath, G. Protein microarrays and proteomics. *Nat. Genet. Suppl.*, 2002, 32: 526–532

Macias, M.J., Wiesner, S. and Sudol, M. WW and SH3 domains, two different scaffolds to recognize proline-rich ligands. *FEBS Letters*, 2002, 513: 30-37

Matuoka, K., Miki, H., Takahashi, K. and Takenawa, T. A novel ligand for an SH3 domain of the adaptor protein Nck bears an SH2 domain and nuclear signaling motifs. *Biochem Biophys Res Comm*, 1997, 239: 488-492

McPherson, P.S., Czernik, A.J., Chilcote, T.J., Onofri, F., Benfenati, F., Greengard, P., Schlessinger, J. and De Camilli, P. Interaction of Grb2 via its Src homology 3 domains with synaptic proteins including synapsin I. *Proc Natl Acad Sci U S A*, 1994, 91, 14: 6486-6490

Mongiovi, A.M., Romano, P.R., Panni, S., Mendoza, M., Wong, W.T., Musacchio, A., Cesareni, G. and Di Fiore, P.P. A novel peptide-SH3 interaction. *EMBO J*, 1999, 18: 5300-5309

Morin, F., Vannier, B., Houdart, F., Regnacq, M., Berges, T. and Voisin, P. A proline-rich domain in the gamma subunit of phosphodiesterase 6 mediates interaction with SH3-containing proteins. *Mol Vis*, 2003, 9: 449-459

Moriki, T., Kuwabara, I., Liu, FT., Maruyama, I.N. Protein domain mapping by lambda phage display: the minimal lactose-binding domain of galectin-3. 1999. *Biochem Biophys Res Commun*., 19, 265, 2: 291-6

Müller, K., Gombert, F.O., Manning, U., Grossmüller, F., Graff, P., Zaegel, H., Zuber, J.F., Freuler, F., et al. Rapid identification of phosphopeptide ligands for SH2 domains. Screening of peptides by fluorescence-activated bead sorting. *J Biol Chem*, 1996, 271: 16500-16505

Niebuhr, K., Ebel, F., Frank, R., Reinhard, M., Domann, E., Carl, U.D., Walter, U., Gertler, F.B., et al. A novel proline-rich motif present in ActA of *Listeria monocytogenes* and cytoskeletal proteins is the ligand for the EVH1 domain, a protein module present in the Ena/VASP family. *EMBO J*, 1997, 16: 5433-5444

Nishizawa, K., Freund, C., Li, J., Wagner, G. and Reinherz, E.L. Identification of a proline-binding motif regulating CD2-triggered T lymphocyte activation. *Proc Natl Acad Sci USA*, 1998, 95: 14897-14902

O'Bryan, J.P., Martin, C.B., Songyang, Z., Cantley, L.C. and Der, C.J. Binding specificity and mutational analysis of the phosphotyrosine binding domain of the brain-specific adaptor protein ShcC. *J Biol Chem*, 1996, 271: 11787-11791

Oligino, L., Lung, F.D., Sastry, L., Bigelow, J., Cao, T., Curran, M., Burke, T.R. Jr, Wang, S., et al. Nonphosphorylated peptide ligands for the Grb2 Src homology 2 domain. *J Biol Chem*, 1997, 272: 29046-29052

Panni, S., Dente, L. and Cesareni, G. *In vitro* evolution of recognition specificity mediated by SH3 domains reveals target recognition rules. *J Biol Chem*, 2002, 277: 21666-21674

Paoluzi, S., Castagnoli, L., Lauro, I., Salcini, A.E., Coda, L., Fre', S., Confalonieri, S., Pelicci, P.G., et al. Recognition specificity of individual EH domains of mammals and yeast. *EMBO J*, 1998, 17: 6541-6550

Pavlik, P., Siegel, R.W., Marzari, R., Sblattero, D., Verzillo, V., Collins, C., Marks, J.D. and Bradbury, A. Predicting antigenic peptides suitable for the selection of phage antibodies. *Hum Antibodies*, 2003, 12, 4: 99-112

Pelicci, G., Dente, L., De Giuseppe, A., Verducci-Galletti, B., Giuli, S., Mele, S., Vetriani, C., Giorgio, M., et al. A family of Shc related proteins with conserved PTB, CH1 and SH2 regions. *Oncogene*, 1996, 13: 633-641

Pero, S.C., Oligino, L., Daly, R.J., Soden, A.L., Liu, C., Roller, P.P., Li, P. and Krag, D.N. Identification of novel non-phosphorylated ligands, which bind selectively to the SH2 domain of Grb7. *J Biol Chem*, 2002, 277: 11918-11926

Phimister, B. Chipping forecast. *Nat Genet*, 1999, 21: 1-60

Pires, J.R., Taha-Nejad, F., Toepert, F., Ast, T., Hoffmüller, U., Schneider-Mergener, J., Kühne, R., Macias, M.J., et al. Solution structures of the YAP65 WW domain and the variant L30K in complex with the peptides GTPPPPYTVG, N-(n-octyl)-GPPPY and PLPPY and the application of peptide libraries reveal a minimal binding epitope. *J Mol Biol*, 2001, 314: 1147-1156

Pirozzi, G., McConnell, S.J., Uveges, A.J., Carter, J.M., Sparks, A.B., Kay, B.K. and Fowlkes, D.M. Identification of novel human WW domain-containing proteins by cloning of ligand targets. *J Biol Chem*, 1997, 272: 14611-14616

Pluthero, F.G. Rapid purification of high-activity Taq DNA polymerase. *Nucleic Acids Res*, 1993, 21, 20: 4850-4851

Prehoda, K.E., Lee, D.J. and Lim, W.A. Structure of the enabled/VASP homology 1 domain-peptide complex: a key component in the spatial control of actin assembly. *Cell*, 1999, 97: 471-480

Ren, R., Ye, Z.S. and Baltimore, D. Abl protein-tyrosine kinase selects the Crk adapter as a substrate using SH3-binding sites. *Genes Dev*, 1994, 8, 7: 783-795

Richter, A., Schwager, C., Hentze, S., Ansorge, W., Hentze, M.W. and Muckenthaler, M. Comparison of fluorescent tag DNA labeling methods used for expression analysis by DNA microarrays. *Biotechniques*, 2002, 33: 620-630

Rickles, R.J., Botfield, M.C., Weng, Z., Taylor, J.A., Green, O.M., Brugge, J.S. and Zoller, M.J. Identification of Src, Fyn, Lyn, PI3K and Abl SH3 domain ligands using phage display libraries. *EMBO J*, 1994, 13: 5598-5604

Rickles, R.J., Botfield, M.C., Zhou, X.M., Henry, P.A., Brugge, J.S. and Zoller, M.J. Phage display selection of ligand residues important for Src homology 3 domain binding specificity. *Proc Natl Acad Sci U S A*, 1995, 92: 10909-10913

Ringstad, N., Nemoto, Y. and De Camilli, P. The SH3p4/Sh3p8/SH3p13 protein family: binding partners for synaptojanin and dynamin via a Grb2-like Src homology 3 domain. *Proc Natl Acad Sci USA*, 1997, 94: 8569-8574

Rivero-Lezcano, O.M., Sameshima, J.H., Marcilla, A. and Robbins, K.C. Physical association between Src homology 3 elements and the protein product of the *c-cbl* proto-oncogene. *J Biol Chem*, 1994, 269: 17363-17366

Rodriguez, M., Li, S.S.-C., Harper, J.W. and Songyang, Z. An Oriented Peptide Array Library (OPAL) strategy to study protein-protein interactions. *J Biol Chem*, 2004, 279: 8802-8807

Rodriguez, M., Yu, X., Chen, J. and Songyang, Z. Phosphopeptide binding specificities of BRCA1 COOH-terminal (BRCT) domains. *J Biol Chem*, 2003, 278: 52914-52918

Rosenberg, A., Griffin, K., Studier, W. S., McCormick, M., Berg, J., Novy, R. and Mierendorf, R. T7 select phage display system: a powerful new protein display system based on bacteriophageT7. 1996, *in Novations*, 6: 1-6

Rost, B. PHD: predicting one-dimensional protein structure by profile-based neural networks. *Methods Enzymol*, 1996, 266: 525–539

Rozakis-Adcock, M., Fernley, R., Wade, J., Pawson, T., Bowtell, D. The SH2 and SH3 domains of mammalian Grb2 couple the EGF receptor to the Ras activator mSos1. *Nature*, 1993, 363, 6424: 15-16

Salcini, A.E., Confalonieri, S., Doria, M., Santolini, E., Tassi, E., Minenkova, O., Cesareni, G., Pelicci, P.G., et al. Binding specificity and in vivo targets of the EH domain, a novel protein-protein interaction module. *Genes Dev*, 1997, 11: 2239-2249

Sambrook, J., Fritsch, E. F. and Maniatis, T. *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, 1989, Plainview, NY

Sambrook, J. and Russel, D.W. Molecular Cloning: A Laboratory Manual (Third Edition) Cold Spring Harbor Laboratory Press, 2001, Cold Spring Harbor, NY

Santamaria, F., Wu, Z., Bolègue, C., Pál, G. and Lu, W. Reexamination of the recognition preference of the specificity pocket of the Abl SH3 domain. *J Mol Recognit,* 2003, 16: 131-138

Santi, E., Capone, S., Mennuni, C., Lahm, A., Tramontano, A., Luzzago, A. and Nicosia, A. Bacteriophage lambda display of complex cDNA libraries: a new approach to functional genomics. *J Mol Biol*, 2000, 296, 2: 497-508

Savitzky, A. and Golay, M.J.E. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Analytical Chemistry*, 1964, 36: 1627–1639

Schmandt, R., Liu, S.K. and McGlade, C.J. Cloning and characterization of mPAL, a novel Shc SH2 domain-binding protein expressed in proliferating cells. *Oncogene*, 1999, 18: 1867-1879

Schmitz, R., Baumann, G. and Gram, H. Catalytic specificity of phosphotyrosine kinases Blk, Lyn, c-Src and Syk as assessed by phage display. *J Mol Biol*, 1996, 260: 664-677

Schneider, S., Buchert, M., Georgiev, O., Catimel, B., Halford, M., Stacker, S.A., Baechi, T., Moelling, K., et al. Mutagenesis and selection of PDZ domains that bind new protein targets. *Nat Biotechnol*, 1999, 17: 170-175

Schumacher, C., Knudsen, B.S., Ohuchi, T., Di Fiore, P.P., Glassman, R.H. and Hanafusa, H. The SH3 domain of Crk binds specifically to a conserved proline-rich motif in Eps15 and Eps15R. *J Biol Chem*, 1995, 270P: 15341-15347

Schultz, J., Hoffmüller, U., Krause, G., Ashurst, J., Macias, M.J., Schmieder, P., Schneider-Mergener, J. and Oschkinat, H. Specific interactions between the syntrophin PDZ domain and voltage-gated sodium channels. *Nat Struct Biol*, 1998, 5: 19-24

Seykora, J.T., Mei, L., Dotto, G.P. and Stein, P.L. Srcasm: a novel Src activating and signaling molecule. *J Biol Chem*, 2002, 277: 2812-2822

Shimomura, Y., Aoki, N., Ito, K. and Ito, M. Gene expression of Sh3d19, a novel adaptor protein with five Src homology 3 domains, in anagen mouse hair follicles. *J Dermatol Sci*, 2003, 31: 43-51

Shrive, A.K., Holden, D., Myles D.A. and Greenhough, T.J. Structure solution of C-reactive proteins: molecular replacement with a twist. *Acta Crystallogr D Biol Crystallogr*, 1996, 52, 6: 1049-1057

Sidhu, S.S., Fairbrother, W.J. and Deshayes, K. Exploring protein-protein interactions with phage display. *ChemBioChem*, 2003, 4, 1, 14-25

Siegel, R.M., Chan, F.K., Chun, H.J. and Lenardo, M.J. The multifaceted role of Fas signaling in immune cell homeostasis and autoimmunity. *Nat Immunol*, 2000, 1, 6: 469-474

Singh, R.R. IL-4 and many roads to lupuslike autoimmunity. *Clin Immunol*, 2003, 108, 2: 73-79.

Smothers, J.F. and Henikoff, S. The HP1 chromo shadow domain binds a consensus peptide pentamer. *Curr Biol*, 2000, 10: 27-30

Soares, M.B., Bonaldo, M.F., Jelene, P., Su, L., Lawton, L. and Efstratiadis A. Construction and characterization of a normalized cDNA library. *Proc Natl Acad Sci USA*, 1994, 91, 20: 9228-3

Sondek, J. and Shortle, D. A general strategy for random insertions and substitutions mutagenesis: substoichiometric coupling of trinucleotide phosphoramidites. *Proc Natl Acad Sci USA,* 1992, 89: 3581-3585

Songyang, Z., Fanning, A.S., Fu, C., Xu, J., Marfatia, S.M., Chishti, A.H., Crompton, A., Chan, A.C., et al. Recognition of unique carboxyl-terminal motifs by distinct PDZ domains. *Science*, 1997, 275: 73-77

Songyang, Z., Margolis, B., Chaudhuri, M., Shoelson, S.E., and Cantley, L.C. The phosphotyrosine interaction domain of SHC recognizes tyrosine-phosphorylated NPXY motif. *J Biol Chem*, 1995, 270: 14863-14866

Songyang, Z., Shoelson, S.E., Chaudhuri, M., Gish, G., Pawson, T., Haser, W.G., King, F., Roberts, T., et al. SH2 domains recognize specific phosphopeptide sequences. *Cell*, 1993, 72: 767-778

Sparks, A.B., Hoffman, N.G., McConnell, S.J., Fowlkes, D.M. and Kay, B.K. Cloning of ligand targets: systematic isolation of SH3 domain-containing proteins. *Nat Biotechnol*, 1996, 14: 741-744

Sparks, A.B., Quilliam, L.A., Thor, J.M., Der, C.J. and Kay, B.K. Identification and characterization of Src SH3 ligands from phage-displayed random peptide libraries. *J Biol Chem*, 1994, 269: 23853-23856

Sternberg, N. and Hoess, R. H. Display of peptides and proteins on the surface of bacteriophage λ. 1995. *Proc. Natl Acad. Sci. USA*, 92: 1609–1613

Stricker, N.L., Christopherson, K.S., Yi, B.A., Schatz, P.J., Raab, R.W., Dawes, G., Bassett, D.E. Jr, Bredt, D.S., et al. PDZ domain of neuronal nitric oxide synthase recognizes novel C-terminal peptide sequences. *Nat Biotechnol*, 1997, 15: 336-342

Tanaka, S., Morishita, T., Hashimoto, Y., Hattori, S., Nakamura, S., Shibuya, M., Matuoka, K., Takenawa, T., et al. C3G, a guanine nucleotide-releasing protein expressed ubiquitously, binds to the Src homology 3 domains of CRK and GRB2/ASH proteins. *Proc Natl Acad Sci USA*, 1994, 91: 3443-3447

Tani, K., Sato, S., Sukezane, T., Kojima, H., Hirose, H., Hanafusa, H. and Shishido, T. Abl interactor 1 promotes tyrosine 296 phosphorylation of mammalian enabled (Mena) by c-Abl kinase. *J Biol Che*, 2003, 278: 21685-21692

Toepert, F., Knaute, T., Guffler, S., Pires, J.R., Matzdorf, T., Oschkinat, H. and Schneider-Mergener, J. Combining SPOT synthesis and native peptide ligation to create large arrays of WW protein domains. *Angew Chem Int Ed*, 2003, 42: 1136-1140

Toepert, F., Pires, J.R., Landgraf, C., Oschkinat, H. and Schneider-Mergener, J. Synthesis of an array comprising 837 variants of the hYAP WW protein domain. *Angew Chem Int Ed*, 2001, 40: 897-900

Tong, A.H.Y., et al., A combined experimental and computational strategy to define protein interaction networks for peptide recognition modules. *Science*, 2002, 295: 321–324

Uetz P, Giot L, Cagney G, Mansfield TA, Judson R.S, et al. A comprehensive analysis of protein–protein interactions in Saccharomyces cerevisiae. *Nature*, 2000, 403: 623–627

Vaccaro, P., Brannetti, B., Montecchi-Palazzi, L., Philipp, S., Helmer-Citterich, M., Cesareni, G. and Dente, L. Distinct binding specificity of the multiple PDZ domains of INADL, a human protein with homology to INAD from *Drosophila melanogaster*. *J Biol Chem*, 2001, 276: 42122-42130

Vidal, M., Braun, P., Chen, E., Boeke, J.D. and Harlow, E. Genetic characterization of a mammalian protein–protein interaction domain by using a yeast reverse two-hybrid system. *Proc Natl Acad Sci USA*, 1996, 93: 10321-10326

Wang, B., Yang, H., Liu, Y.C., Jelinek, T., Zhang, L., Ruoslahti, E. and Fu, H. Isolation of high-affinity peptide antagonists of 14-3-3 proteins by phage display. *Biochemistry*, 1999, 38: 12499-12504

Weng, Z., Thomas, S.M., Rickles, R.J., Taylor, J.A., Brauer, A.W., Seidel-Dugan, C., Michael, W.M., Dreyfuss, G. and Grugge, J.S. Identification of Src, Fyn and Lyn SH3-binding proteins: implications for a function of SH3 domains. *Mol Cell Biol*, 1994, 14: 4509– 4521

Weng, Z., Rickles, R.J., Feng, S., Richard, S., Shaw, A.S., Schreiber, S.L. and Brugge, J.S. Structure-function analysis of SH3 domains: SH3 binding specificity altered by single amino acid substitutions. *Mol Cell Biol*, 1995, 15, 10: 5627-5634

Weng, Z., Taylor, J. A., Turner, C.E., Brugge, J. S. and Seidel-Dugan, C. Detection of Src homology 3-binding proteins, including paxillin, in normaland v-Src–transformed Balb/c 3T3 cells. *J. Biol. Chem*, 1993, 268: 14956–14963

Wong, H.C, Bourdelas, A., Krauss, A., Lee, H.J, Shao, Y., Wu, D., Mlodzik, M., Shi, D.L. and Zheng, J. Direct binding of the PDZ domain of Dishevelled to a conserved internal sequence in the C-terminal region of Frizzled. *Mol Cell*, 2003, 5: 1251-1260

Wood, J.D., Yuan, J., Margolis, R.L., Colomer, V., Duan, K., Kushi, J., Kaminsky, Z., Kleiderlein, J.J. Jr., et al. Atrophin-1, the DRPLA gene product, interacts with two families of WW domain-containing proteins. *Mol Cell Neurosci*, 1998, 11: 149-160

Yaffe, M.B., Rittinger, K., Volinia, S., Caron, P.R., Aitken, A., Leffers, H., Gamblin, S.J., Smerdon, S.J., et al. The structural basis for 14-3-3: phosphopeptide binding specificty. *Cell*, 1997, 91: 961-971

Yamabhai, M., Hoffman, N.G., Hardison, N.L., McPherson, P.S., Castagnoli, L., Cesareni, G. and Kay, B.K. Intersectin, a novel adaptor protein with two Eps15 homology and five Src homology 3 domains. *J Biol Chem*, 1998, 273: 31401-31407

Ymer, S. Heat inactivation of DNA ligase prior to electroporation increases transformation efficiency. *Nucl. Acids Res*., 1991, 19: 6960

Yu, H., Chen, J.K., Feng, S., Dalgarno, D.C., Brauer, A.W. and Schreiber, S.L. Structural basis for the binding of proline-rich peptides to SH3 domains. *Cell,* 1994, 76: 933-945

Yu, J., Othman, M.I., Farjo, R., Zareparsi, S., MacNee, S.P., Yoshida, S. and Swaroop, A. Evaluation and optimization of procedures for target labeling and hybridization of cDNA microarrays. *Mol Vis*, 2002, 8: 130-137

Zhou, D., Quach, K.M., Yang, C., Lee, S.Y., Pohajdak, B. and Chen, S. PNRC: a proline-rich nuclear receptor coregulatory protein that modulates transcriptional activation of multiple nuclear receptors including orphan receptors SF1 (steroidogenic factor 1) and ERRα1 (estrogen related receptor α-1). *Mol Endocrinol*, 2000, 14: 986-998

Zhu, H., Bilgin, M., Bangham, R., Hall, D., Casamayor, A., Bertone, P., Lan, N., Jansen, R., et al. Global analysis of protein activities using proteome chips. *Science*, 2001, 293: 2101-2105

Zhu, J. and Shore, S.K. c-ABL tyrosine kinase activity is regulated by association with a novel SH3-domain-binding protein. *Mol Cell Biol*, 1996, 16: 7054-7062

Ziemnicka-Kotula, D., Xu, J., Gu, H., Potempska, A., Kim, K.S., Jenkins, E.C., Trenkner, E. and Kotula, L. Identification of a candidate human spectrin Src homology 3 domain-binding protein suggests a general mechanism of association of tyrosine kinases with the spectrin-based membrane skeleton. *J Biol Chem*, 1998, 273

Zozulya, S., Lioubin, M., Hill, R. J., Abram, C., Gishizky, M. L., Mapping signal transduction pathways by phage display. *Nat. Biotechnol*.,1999, 17: 1193–1198

Zucconi, A., Dente, L., Santonico, E., Castagnoli, L. and Cesareni, G. Selection of ligands by panning of domain libraries displayed on phage lambda reveals new potential partners of synaptojanin 1. *J Mol Biol*, 2001, 307: 1329-1339

6. Appendix

Appendix 1

Parameters used for printing the amygdala biochips using Micro Grid II mictoarrayer (Biorobotics).

[Parameters]
RunType=MicroArray
BarcodesEnabled=0
HaltBarcodes=1
CacheBarcodes=0
ClearBarcodes=0
WhereBarcodes=1
ExpectedBarcode1=[Unknown]
ExpectedBarcode2=
ExpectedBarcode3=
ExpectedBarcode4=
ExpectedBarcode5=
ExpectedBarcode6=
ExpectedBarcode7=
ExpectedBarcode8=
ExpectedBarcode9=
ExpectedBarcode10=
ExpectedBarcode11=
ExpectedBarcode12=
ExpectedBarcode13=
ExpectedBarcode14=
ExpectedBarcode15=
ExpectedBarcode16=
ExpectedBarcode17=
ReloadPins=1
NumSourcePlates=17
WashAfterEveryVisit=1
WashAfterEverySample=0
RequestLogInfo=0
SourcePlate=384standardb.microplate
ToolType=4x4 384 split.tool
PromptPlates=0
PromptPlatesInterval=1
SourceWiggle=0
SourceDibble=0
RemoveOneLid=1
TargetDip=0
TargetWiggle=0
SoftTouchDistance=0.300000011920929
SoftTouchSpeed=4
DelayBefore=0
TargetDwell=0
DelayAfterSpotting=0
MultiStrikes=1
SoftTouch=0
TargetDepth=0.300000011920929
ILayout=0
PrespotRecycle=0
UseCamera=0
Microplates=1
PreSpot=1
SlideSize=25
NumSamples=391
LastPlateVisits=7
SpotsPerSource=57
AdapterPlate=slidesvertical extra.adapter
AdapterPlate2=slidesvertical extra.adapter
[PinGrid]
XSpots=20
YSpots=20
Line0=385, 386, 387, 388, 389, 390, 134, 154, 174, 194, 214, 234, 254, 274, 294, 314, 334, 354, 374, 0,
Line1=19, 38, 57, 76, 95, 114, 133, 153, 173, 193, 213, 233, 253, 273, 293, 313, 333, 353, 373, 0,
Line2=18, 37, 56, 75, 94, 113, 132, 152, 172, 192, 212, 232, 252, 272, 292, 312, 332, 352, 372, 0,

Line3=17, 36, 55, 74, 93, 112, 131, 151, 171, 191, 211, 231, 251, 271, 291, 311, 331, 351, 371, 391,
Line4=16, 35, 54, 73, 92, 111, 130, 150, 170, 190, 210, 230, 250, 270, 290, 310, 330, 350, 370, 390,
Line5=15, 34, 53, 72, 91, 110, 129, 149, 169, 189, 209, 229, 249, 269, 289, 309, 329, 349, 369, 389,
Line6=14, 33, 52, 71, 90, 109, 128, 148, 168, 188, 208, 228, 248, 268, 288, 308, 328, 348, 368, 388,
Line7=13, 32, 51, 70, 89, 108, 127, 147, 167, 187, 207, 227, 247, 267, 287, 307, 327, 347, 367, 387,

Line8=12, 31, 50, 69, 88, 107, 126, 146, 166, 186, 206, 226, 246, 266, 286, 306, 326, 346, 366, 386,
Line9=11, 30, 49, 68, 87, 106, 125, 145, 165, 185, 205, 225, 245, 265, 285, 305, 325, 345, 365, 385,
Line10=10, 29, 48, 67, 86, 105, 124, 144, 164, 184, 204, 224, 244, 264, 284, 304, 324, 344, 364, 384,
Line11=9, 28, 47, 66, 85, 104, 123, 143, 163, 183, 203, 223, 243, 263, 283, 303, 323, 343, 363, 383,
Line12=8, 27, 46, 65, 84, 103, 122, 142, 162, 182, 202, 222, 242, 262, 282, 302, 322, 342, 362, 382,
Line13=7, 26, 45, 64, 83, 102, 121, 141, 161, 181, 201, 221, 241, 261, 281, 301, 321, 341, 361, 381,
Line14=6, 25, 44, 63, 82, 101, 120, 140, 160, 180, 200, 220, 240, 260, 280, 300, 320, 340, 360, 380,
Line15=5, 24, 43, 62, 81, 100, 119, 139, 159, 179, 199, 219, 239, 259, 279, 299, 319, 339, 359, 379,
Line16=4, 23, 42, 61, 80, 99, 118, 138, 158, 178, 198, 218, 238, 258, 278, 298, 318, 338, 358, 378,
Line17=3, 22, 41, 60, 79, 98, 117, 137, 157, 177, 197, 217, 237, 257, 277, 297, 317, 337, 357, 377,
Line18=2, 21, 40, 59, 78, 97, 116, 136, 156, 176, 196, 216, 236, 256, 276, 296, 316, 336, 356, 376,
Line19=1, 20, 39, 58, 77, 96, 115, 135, 155, 175, 195, 215, 235, 255, 275, 295, 315, 335, 355, 375,
NthOrder=1
PatternFormat=2
ToolGridIndex=1, 1,
Splice=0
Pitch=0.215000003576279
NumCopies=58
[SlideLayout]
BottomMargin=19.7000007629395
CentreHoriz=0
CentreVert=0
TopMargin=0.519999980926514
LeftMargin=3.45000004768372
RightMargin=3.45000004768372
XSpacing=2
YSpacing=2
[Wash]
InterCycleDelay=5.5
LeftSeconds=2
RightSeconds=2
DrySeconds=2
VacuumEnabled=1
Cycles=2
MWSCycles=3
[RunPrefs]
PrintReplicatesFirst=0
SourceDipTop=0
SourceDipBottom=0
TargetDipTop=0
TargetDipBottom=0
VisionID=
BioBankID=DEFAULT
IOControl=0
Bath12ZOffset=0
Bath3ZOffset=0
FeatureOffsetX=0
FeatureOffsetY=0
HumidStartRate=63
HumidRunRate=58
TargetHumidity=126
MinHumidity=99
UseHumidityControl=0
UseHumidityHEPA=0
PauseHumidity=0
EndHumidityOn=0
PrimeBath=1
AskLoadTool=1
WashAtStart=1
WashAtEnd=1
AskBioBank=1
AskTrays=1
SourceDwellTime=0.5
TargetDwellTime=1
SourceWiggles=0
SourceWiggleSize=0.25
SourceDips=1
SourceDipSize=2
TargetWiggles=0
TargetWiggleSize=0.25

TargetDips=0
TargetDipSize=2
SourceDepth=-0.5
TargetDepth=0
BathFillTime=2
BathDrainTime=2.5
RecirculateAction=1
RecirculateStatic=0
RecirculateWiggleSize=2
RecirculateDipSize=1
BlotFirstTimes=1
BlotFirstSlides=10
CacheBarcodes=1
HaltNoWater=1
BathWiggles=4
BathWiggleSize=0
HeaterOnTime=0
HeaterOffTime=0
HeaterCoolTime=0
SourceSoftTouchSpeed=4
TargetSoftTouchSpeed=4
SourceSoftTouchEnabled=1
TargetSoftTouchEnabled=0
SourceSoftTouchDistance=8
TargetSoftTouchDistance=1
[PreSpotLayout]
BottomMargin=1.41999995708466
CentreHoriz=0
CentreVert=0
TopMargin=0.990000009536743
LeftMargin=3.40000009536743
RightMargin=3.40000009536743
XSpacing=0
YSpacing=0
ToolGridIndex=1, 2, 3, 4,
Splice=0
[PreSpotting]
Pitch=0.300000011920929
SoftTouchDistance=0.300000011920929
SoftTouchSpeed=4
DelayBefore=0
TargetDwell=0
DelayAfterSpotting=0
MultiStrikes=1
SoftTouch=0
TargetDepth=0.300000011920929
ISoftTouchDistance=0
ISoftTouchSpeed=0
IPitch=0
IDelayBefore=1
ITargetDwell=1
IDelayAfterSpotting=1
IMultiStrikes=1
ISoftTouch=0
ITargetDepth=1
[LogFile]
Operator_name*=_
Library=_
Gridding_Number=_
Set_Number=_
Replica_Number=_
Date=_
Time=_

Appendix 2

List of 98 human proteins and their mouse homologs, as well as 14 KIAA proteins, chosen for peptide preparation using ASS software and chemical synthesis on cellulose membranes.

Mouse	Human	Mouse	Human	Mouse	Human
gi|10946574	gi|17939432	gi|19527234	gi|12052851	gi|6685763	gi|19264127
gi|12018296	gi|19684120	gi|19527238	gi|13676377	gi|6685997	gi|3386485
gi|12746424	gi|12804042	gi|19527258	gi|7160984	gi|20867606	gi|15559431
gi|12805413	gi|14286219	gi|19527306	gi|1906008	gi|20873704	gi|183369
gi|12805431	gi|12653340	gi|20341045	gi|563885	gi|20877156	gi|12654186
gi|12844082	gi|14705275	gi|20341854	gi|12653396	gi|21311919	gi|14424792
gi|12845889	gi|13278740	gi|20343963	gi|2065176	gi|21312004	gi|297901
gi|12849017	gi|9963840	gi|20820655	gi|14250498	gi|6753050	gi|19584409
gi|12850132	gi|12803908	gi|20821322	gi|20521841	gi|6753676	gi|2967518
gi|12858691	gi|14250078	gi|20830475	gi|12653600	gi|6754974	gi|20521899
gi|12963653	gi|12804092	gi|20832179	gi|13477172	gi|6755682	gi|16356636
gi|12963697	gi|12052958	gi|20833478	gi|1869811	gi|6755714	gi|2335046
gi|13384704	gi|15559275	gi|20833765	gi|19913486	gi|6755995	gi|12803340
gi|13384746	gi|285902	gi|20837251	gi|15277483	gi|6756039	gi|23221
gi|13384810	gi|23274139	gi|20840384	gi|18043711	gi|7022124	gi|7022123
gi|13385500	gi|13938546	gi|20844633	gi|7020550	gi|7305061	gi|18089153
gi|13386156	gi|183300	gi|20863245|	gi|16307524	gi|7656991	gi|5701730
gi|13592137	gi|24660071	gi|20880590	gi|12654886	gi|7710102	gi|6934239
gi|13637964	gi|29635	gi|20882495	gi|14250348	gi|8217551	gi|2921407
gi|13905144	gi|23272131	gi|20889057	gi|14279193	gi|9790219	gi|415586
gi|13929082	gi|1946348	gi|20901301	gi|3378167	**KIAA**	
gi|14043149	gi|14043148	gi|20909498	gi|1504033	gi|20520995	
gi|1407651	gi|2407912	gi|20909548	gi|21755897	gi|20520999	
gi|14777622	gi|182190	gi|20910093	gi|2282031	gi|20521017	
gi|15990388	gi|17391416	gi|20912095	gi|19070131	gi|20521013	
gi|16758814|	gi|1809264	gi|20912830	gi|312823	gi|20521031	
gi|1730217	gi|2570403	gi|20978553	gi|13278911	gi|20521051	
gi|17390379	gi|15214594	gi|23879	gi|17389605	gi|20521039	
gi|17390865	gi|15082295	gi|2497503	gi|1463025	gi|20521041	
gi|17865671	gi|13529265	gi|3122018	gi|2967518	gi|20521045	
gi|18201867	gi|18448714	gi|3122037	gi|2342485	gi|20521049	
gi|18202285	gi|31107	gi|346651	gi|21104443	gi|20521067	
gi|18490449	gi|15866713	gi|3810884	gi|1944339	gi|20521069	
gi|18702315	gi|1463025	gi|487855	gi|181848	gi|20521071	
gi|14198445	gi|12653166	gi|6679140	gi|21040483	gi|20521075	
gi|19424252	gi|3281993	gi|6679573	gi|190749		
gi|19526884	gi|13938606	gi|6680137	gi|1513314		
gi|19526926	gi|12052735	gi|6680908	gi|13477282		
gi|19527010	gi|13543435	gi|6680924	gi|15126675		

Appendix 3

List of peptides from membrane No.1, used for high-throughout panning with T7 phage display library as well as average amounts of phages individually eluted from each peptide-spot.

Plate position	Peptide sequence	Gi code	Amounts of eluted phages
PL1 A1	SAHNNHMAKVLTPELYAELRAKSTP	**gi17939432**	1.41E+06
PL1 A2	GDEESYEVFKDLFDPIIEDRHGGYK		3.05E+06
PL1 A3	IIEDRHGGYKPSDEHKTDLNPDNLQ		3.09E+06
PL1 A4	KTDLNPDNLQGGDDLDPNYVLSSRV		2.78E+06
PL1 A5	RVRTGRSIRGFCLPPHCSRGERRAI		5.28E+06
PL1 A6	YYALKSMTEAEQQQLIDDHFLFDKP		1.44E+06
PL1 A7	MARDWPDARGIWHNDNKTFLVWVNE		2.25E+06
PL1 A8	NKTFLVWVNEEDHLRVISMQKGGNM		2.87E+06
PL1 A9	FCTGLTQIETLFKSKDYEFMWNPHL		8.40E+05
PL1 A10	KLPNLGKHEKFSEVLKRLRLQKRGT		5.42E+06
PL1 A11	VVDGVKLLIEMEQRLEQGQAIDDLM		1.08E+06
PL1 A12	YIDSKFEDYLNAESRVNRRQMPDNR	**gi19684120**	3.00E+05
PL1 B1	QEHKIKIYEFPETDDEEENKLVKKI		6.45E+05
PL1 B2	HYENYRSRKLAAVTYNGVDNNKNKG		2.01E+06
PL1 B3	VDNNKNKGQLTKSPLAQMEEERREH		1.22E+06
PL1 B4	QKLKDSEAELQRRHEQMKKNLEAQH		7.50E+05
PL1 B5	QMKKNLEAQHKELEEKRRQFEDEKA		7.05E+05
PL1 B6	ANWEAQQRILEQQNSSRTLEKNKKK		6.75E+05
PL1 B7	VLPDKETSLVDAYEKCRGLADPKVC	**gi12804042**	1.71E+06
PL1 B8	REKGVNSFQMFMTYKDLYMLRDSEL		2.39E+06
PL1 B9	GIEISRPEELEAEATHRVITIANRT		1.31E+06
PL1 B10	ETTTAHATLTGLHYYHQDWSHAAAY		2.09E+06
PL1 B11	HQDWSHAAAYVTVPPLRLDTNTSTY		2.63E+06
PL1 B12	ASDHRPFTTKQKAMGKEDFTKIPHG		3.14E+06
PL1 C1	GKMDENRFVAVTSSNAAKLLNLYPR		1.04E+06
PL1 C2	NLYPRKGRIIPGADADVVVWDPEAT		1.13E+06
PL1 C3	RSFPDTVYKKLVQREKTLKVRGVDR		7.05E+06
PL1 C4	KVRGVDRTPYLGDVAVVVHPGKKEM		3.14E+06
PL1 C5	EMGTPLADTPTRPVTRHGGMRDLHE		4.80E+05
PL1 C6	MRDLHESSFSLSGSQIDDHVPKRAS		2.10E+05
PL1 C7	FEYIIAEKRGKNNTVGLIQLNRPKA	**gi14286219**	6.45E+05
PL1 C8	MQNLSFQDCYSSKFLKHWDHLTQVK		1.80E+05
PL1 C9	GEKAQFAQPEILIGTIPGAGGTQRL		7.50E+04
PL1 C10	TRAVGKSLAMEMVLTGDRISAQDAK		1.50E+05
PL1 C11	AFEMTLTEGSKLEKKLFYSTFATDD		1.20E+05
PL1 C12	LFYSTFATDDRKEGMTAFVEKRKAN		3.90E+05
PL1 D1	LVRRPLREVSGLLKRRFHWTAPAAL	**gi12653340**	4.74E+06
PL1 D2	NQGMDEELERDEKVFLLGEEVAQYD		3.60E+05
PL1 D3	EEVAQYDGAYKVSRGLWKKYGDKRI		6.90E+05
PL1 D4	SAAKTYYMSGGLQPVPIVFRGPNGA		4.05E+05
PL1 D5	LKVVSPWNSEDAKGLIKSAIRDNNP		2.10E+05
PL1 D6	FEFPPEAQSKDFLIPIGKAKIERQG		1.50E+05
PL1 D7	RGAGQQQSQEMMEVDRRVESEESGD	**gi14705275**	1.50E+05
PL1 D8	RRVESEESGDEEGKKHSSGIVADLS		2.10E+05
PL1 D9	ADLSEQSLKDGEERGEEDPEEEHEL		4.50E+04

Plate position	Peptide sequence	Gi code	Amounts of eluted phages
PL1 D10	EEDPEEEHELPVDMETINLDRDAED		4.20E+05
PL1 D11	NLEELQSLRELDLYDNQIKKIENLE		5.40E+05
PL1 D12	GLENNNKLTMLDIASNRIKKIENIS		1.49E+06
PL1 E1	LKGARSLETVYLERNPLQKDPQYRR		1.61E+06
PL1 E2	ADSRATAGAYIASQTVKKVIEINPY	**gi13278740**	6.60E+05
PL1 E3	YELRNKERISVAAASKLLANMVYQY		1.50E+06
PL1 E4	GWDKRGPGLYYVDSEGNRISGATFS		1.95E+05
PL1 E5	MDRGYSYDLEVEQAYDLARRAIYQA		1.65E+05
PL1 E6	DLARRAIYQATYRDAYSGGAVNLYH		7.50E+04
PL1 E7	LYHVREDGWIRVSSDNVADLHEKYS		6.00E+04
PL1 E8	RNPRKKAAKELRLQGAEVVQGDQDD	**gi9963840**	6.00E+06
PL1 E9	NYWESCSQEQEVKQGKLLADLARRL		1.50E+04
PL1 E10	NIKKLTAGRLAAAHFDGKGEVEEYF		1.31E+06
PL1 E11	HFLPQKAPDGKSYLLSLPTGDVPMD		1.65E+05
PL1 E12	LKMPEKYVGQNIGLSTCRHTAEEYA		1.35E+05
PL1 F1	TKHTRKVVHDAKMTPEDYEKLGFPG		1.86E+06
PL1 F2	DLANMFRFYALRPDRDIELTLRLNP		1.65E+05
PL1 F3	DIELTLRLNPKALTLDQWLEQHKGD		9.00E+04
PL1 F4	AERQEALREFVAVTGAEEDRARFFL	**gi12803908**	2.55E+05
PL1 F5	LIHDQDEDEEEEEGQRFYAGGSERS		4.50E+04
PL1 F6	RFYAGGSERSGQQIVGPPRKKSPNE		1.76E+06
PL1 F7	GAKEHGAVAVERVTKSPGETSKPRP		1.95E+05
PL1 F8	YRLGAAPEEESAYVAGEKRQHSSQD		2.55E+05
PL1 F9	DNGELRSYQDPSNAQFLESIRRGEV		1.80E+05
PL1 F10	DMEDHRDEDFVKPKGAFKAFTGEGQ		3.15E+05
PL1 F11	SPAQQAENEAKASSSILIDESEPTT		3.30E+05
PL1 F12	ETVADTRRLITKPQNLNDAYGPPSN	**gi14250078**	1.65E+05
PL1 G1	KLKESTVRRRYSDFEWLRSELERES		1.46E+06
PL1 G2	LPFRGDDGIFDDNFIEERKQGLEQF		6.15E+05
PL1 G3	LAQNERCLHMFLQDEIIDKSYTPSK		3.00E+05
PL1 G4	DEEKLPPGWEKRMSRSSGRVYYFNH		5.55E+05
PL1 G5	TNASQWERPSGNSSSGGKNGQGEPA		3.00E+04
PL1 G6	LVKHSQSRRPSSWRQEKITRTKEEA		3.71E+06
PL1 G7	SRFWEWGKNIVCVGRNYADHVREMR	**gi12052958**	2.25E+05
PL1 G8	KPSTAYAPEGSPILMPAYTRNLHHE		2.25E+05
PL1 G9	DMTARDVQDECKKKGLPWTLAKSFT		3.90E+05
PL1 G10	SFTASCPVSAFVPKEKIPDPHKLKL		2.55E+05
PL1 G11	KIPDPHKLKLWLKVNGELRQEGETS		1.05E+06
PL1 G12	GPVKENDEIEAGIHGLVSMTFKVEK		1.35E+05
PL1 H1	SRPPPPRVASVLGTMEMGRRMDAPA	**gi15559275**	2.10E+05
PL1 H2	RAFLERGHTELDTAFMYSDGQSETI		3.30E+05
PL1 H3	ATKANPWDGKSLKPDSVRSQLETSL		2.10E+05
PL1 H4	KSNGWILPTVYQGMYNATTRQVETE		1.05E+05
PL1 H5	LLTGKYKYEDKDGKQPVGRFFGNSW		5.85E+05
PL1 H6	PVGRFFGNSWAETYRNRFWKEHHFE		1.50E+04
PL1 H7	SLEQLEQNLAATEEGPLEPAVVDAF		2.70E+05
PL1 H8	TSKGPEEEHPSVTLFRQYLRIRTVQ	**gi285902**	1.80E+05
PL1 H9	QYLRIRTVQPKPDYGAAVAFFEETA		4.20E+05
PL1 H10	KEHWSHDPFEAFKDSEGYIYARGAQ		3.90E+05
PL1 H11	FVPDEEVGGHQGMELFVQRPEFHAL		4.50E+05
PL1 H12	FYSERSPWWVRVTSTGRPGHASRFM		5.55E+05
PL2 A1	AFREKEWQRLQSNPHLKEGSVTSVN		4.95E+05
PL2 A2	ATMSASFDFRVAPDVDFKAFEEQLQ		9.00E+04

Plate position	Peptide sequence	Gi code	Amounts of eluted phages
PL2 A3	EGVTLEFAQKWMHPQVTPTDDSNPW		2.55E+05
PL2 A4	AATDNRYIRAVGVPALGFSPMNRTP		1.35E+05
PL2 A5	LGFSPMNRTPVLLHDHDERLHEAVF		1.50E+05
PL2 A6	FRKHTQAEGKKLGLVGWVQNTDRGT	**gi23274139**	1.98E+06
PL2 A7	VRHMQEWLETRGSPKSHIDKANFNN		7.20E+05
PL2 A8	QGSSDPEEESVLYSNRAACHLKDGN	**gi13938546**	9.00E+04
PL2 A9	VSAQKRWNSLPSENHKEMAKSKSKE		9.15E+05
PL2 A10	KEMAKSKSKETTATKNRVPSAGDVE		2.25E+05
PL2 A11	ARVLKEEGNELVKKGNHKKAIEKYS		2.25E+05
PL2 A12	LKQYTEAVKDCTEALKLDGKNVKAF		1.50E+04
PL2 B1	LDGKNVKAFYRRAQAHKALKDYKSS		1.85E+06
PL2 B2	FADISNLLQIEPRNGPAQKLRQEVK		1.35E+05
PL2 B3	LEYTDSSYEEKKYTMGDAPDYDRSQ	**gi183300**	8.10E+05
PL2 B4	GDAPDYDRSQWLNEKFKLGLDFPNL		3.00E+04
PL2 B5	CGESEKEQIREDILENQFMDSRMQL		4.50E+04
PL2 B6	QFMDSRMQLAKLCYDPDFEKLKPEY		1.50E+04
PL2 B7	SSARARRSLFASTAQSPRILRSEPV	**gi24660071**	2.57E+06
PL2 B8	SPRILRSEPVRTPPIPAFSPLKTLR		1.08E+06
PL2 B9	AWNRDRTQIALSPNNHEVHIYKKNG		2.10E+05
PL2 B10	QWVKAHELKEHNGHITGIDWAPKSD		6.00E+04
PL2 B11	FESENDWWVSKHIKKPIRSTVLSLD		3.00E+05
PL2 B12	AYIKEVDEKPASTPWGSKMPFGQLM		2.40E+05
PL2 C1	NYDDRGCLTFVSKLDIPKQSIQRNM		2.10E+05
PL2 C2	QRNMSAMERFRNMDKRATTEDRNTA		2.04E+06
PL2 C3	NTALETLHQNSITQVSIYEVDKQDC		2.40E+05
PL2 C4	GPQQQPPYLHLAELTASQFLEIWKH	**gi29635**	1.65E+05
PL2 C5	NFFQELEKARKGSGMMSKSDNFGEK		1.31E+06
PL2 C6	MSKSDNFGEKMKEFMQKYDKNSDGK		4.05E+05
PL2 C7	SSAEFMEAWRKYDTDRSGYIEANEL		2.70E+05
PL2 C8	LLKKANRPYDEPKLQEYTQTILRMF		1.83E+06
PL2 C9	FYDKDRSGYIDEHELDALLKDLYEK		2.40E+05
PL2 C10	ALLKDLYEKNKKEINIQQLTNYRKS		1.41E+06
PL2 C11	IQQLTNYRKSVMSLAEAGKLYRKDL		1.10E+06
PL2 C12	ILCPRRTTAQLGPRRNPAWSLQAGR	**gi23272131**	1.67E+06
PL2 D1	NPAWSLQAGRLFSTQTAEDKEEPLH		1.20E+05
PL2 D2	GSTSKHEFQAETKKLLDIVARSLYS		1.20E+05
PL2 D3	SLYSEKEVFIRELISNASDALEKLR		9.00E+04
PL2 D4	LRHKLVSDGQALPEMEIHLQTNAEK		3.45E+05
PL2 D5	MMDPKDVGEWQHEEFYRYVAQAHDK		6.00E+04
PL2 D6	YRYVAQAHDKPRYTLHYKTDAPLNI		1.20E+05
PL2 D7	VLQQRLIKFFIDQSKKDAEKYAKFF		4.20E+05
PL2 D8	TATEQEVKEDIAKLLRYESSALPSG		1.20E+05
PL2 D9	YLCAPNRHLAEHSPYYEAMKKKDTE		1.05E+05
PL2 D10	REFDKKKLISVETDIVVDHYKEEKF		1.65E+05
PL2 D11	DHYKEEKFEDRSPAAECLSEKETEE		1.20E+05
PL2 D12	RMQQLAKTQEERAQLLQPTLEINPR		3.00E+05
PL2 E1	QVLNSDELQELYEGLRLNNMNKYDY	**gi1946348**	1.65E+05
PL2 E2	TGYTRDKSFLAMVVDIVQELKQQNP		1.65E+05
PL2 E3	VLGDKWDGEGSMYVPEDLLPVYKEK		4.05E+05
PL2 E4	SSDLPSPQGSNYLIVLGSQRRRNPA		3.15E+05
PL2 E5	WTHKHPNNLKVACEKTVSTLHHVLQ		1.13E+06
PL2 E6	GEGVRPSPMQLELRMVQSKRDIEDP		9.15E+05
PL2 E7	EIMSKLAGESESNLRKAFEEAEKNA	**gi14043148**	1.50E+05

Plate position	Peptide sequence	Gi code	Amounts of eluted phages
PL2 E8	AEKNAPAIIFIDELDAIAPKREKTH		6.00E+04
PL2 E9	ATNRPNSIDPALRRFGRFDREVDIG		2.78E+06
PL2 E10	DDFRWALSQSNPSALRETVVEVPQV		4.50E+04
PL2 E11	EDVKRELQELVQYPVEHPDKFLKFG		2.25E+05
PL2 E12	GPELLTMWFGESEANVREIFDKARQ		4.50E+04
PL2 F1	ILTEMDGMSTKKNVFIIGATNRPDI		1.20E+05
PL2 F2	PLPDEKSRVAILKANLRKSPVAKDV		7.05E+05
PL2 F3	ESIESEIRRERERQTNPSAMEVEED		1.50E+04
PL2 F4	NPSAMEVEEDDPVPEIRRDHFEEAM		3.00E+04
PL2 F5	RRDHFEEAMRFARRSVSDNDIRKYE		1.97E+06
PL2 F6	QGGAGPSQGSGGGTGGSVYTEDNDD		1.50E+05
PL2 F7	TLNMKNYKGYEKKPYCNAHYPKQSF	**gi2407912**	4.35E+05
PL2 F8	ENLRLKQQSELQSQVRYKEEFEKNK		7.50E+04
PL2 F9	PELQRIKKTQDQISNIKYHEEFEKS		1.80E+05
PL2 F10	EGMEPERRDSQDGSSYRRPLEQQQP		1.50E+04
PL2 F11	YRRPLEQQQPHHIPTSAPVYQQPQQ		1.20E+05
PL2 F12	VGPRYTNLSYIGEGAYGMVCSAYDN	**gi182190**	1.20E+05
PL2 G1	ISPFEHQTYCQRTLREIKILLRFRH		2.55E+05
PL2 G2	QILRGLKYIHSANVLHRDLKPSNLL		7.05E+05
PL2 G3	ARVADPDHDHTGFLTEYVATRWYRA		1.35E+05
PL2 G4	EYVATRWYRAPEIMLNSKGYTKSID		4.95E+05
PL2 G5	SPSQEDLNCIINLKARNYLLSLPHK		2.70E+05
PL2 G6	RNYLLSLPHKNKVPWNRLFPNADSK		1.44E+06
PL2 G7	LDKMLTFNPHKRIEVEQALAHPYLE		1.20E+05
PL2 G8	EQALAHPYLEQYYDPSDEPIAEAPF		2.70E+05
PL2 G9	SDEPIAEAPFKFDMELDDLPKEKLK		1.35E+05
PL2 G10	LARRKPVLPALTINPTIAEGPSPTS	**gi17391416**	9.45E+05
PL2 G11	VDLQKKLEELELDEQQKKRLEAFLT		1.50E+05
PL2 G12	QKKRLEAFLTQKAKVGELKDDDFER		1.68E+06
PL2 H1	DDDFERISELGAGNGGVVTKVQHRP		9.00E+04
PL2 H2	EHMDGGSLDQVLKEAKRIPEEILGK		1.05E+05
PL2 H3	YLREKHQIMHRDVKPSNILVNSRGE		5.25E+05
PL2 H4	RYPIPPPDAKELEAIFGRPVVDGEE		6.00E+04
PL2 H5	PVVDGEEGEPHSISPRPRPPGRPVS		1.80E+05
PL2 H6	LLDYIVNEPPPKLPNGVFTPDFQEF		7.50E+04
PL2 H7	KNPAERADLKMLTNHTFIKRSEVEE		2.70E+05
PL2 H8	SEVEEVDFAGWLCKTLRLNQPGTPT		7.50E+04
PL2 H9	TIPRDIDKQYVGFATLPNQVHRKSV	**gi1809264**	9.00E+04
PL2 H10	TDLYKDRKLLSAEERISQTVEILKH		1.05E+05
PL2 H11	DYVDQQFEQYFRDESGLNRKNIQDN		3.45E+05
PL2 H12	SEIRKLKERIREEIDKFGIHVYQFP		3.15E+05
PL3 A1	YQFPECDSDEDEDFKQQDRELKESA		3.45E+05
PL3 A2	HYENYRAHCIQQMTSKLTQDSRMES		4.05E+05
PL3 A3	SPIPILPLPTPDAETEKLIRMKDEE		2.40E+05
PL3 A4	EKLIRMKDEELRRMQEMLQRMKQQM		1.01E+06
PL3 A5	NETLASLKSEAESLKGKLEEERAKL	**gi2570403**	5.40E+05
PL3 A6	MKTRRTLKGHGNKVLCMDWCKDKRR		1.30E+07
PL3 A7	KRRIVSSSQDGKVIVWDSFTTNKEH		5.25E+05
PL3 A8	YPLTFDKNENMAAKKKSVAMHTNYL		1.04E+06
PL3 A9	QAFETHESDINSVRYYPSGDAFASG		1.35E+05
PL3 A10	YPSGDAFASGSDDATCRLYDLRADR		4.50E+04
PL3 A11	GYNDYTINVWDVLKGSRVSILFGHE		2.85E+05
PL3 A12	YLGSRQKHSLPDLPYDYGALEPHIN	**gi15214594**	7.50E+04

Plate position	Peptide sequence	Gi code	Amounts of eluted phages
PL3 B1	LHHSKHHAAYVNNLNVTEEKYQEAL		4.95E+05
PL3 B2	GEPKGELLEAIKRDFGSFDKFKEKL		1.35E+05
PL3 B3	GFNKERGHLQIAACPNQDPLQGTTG		2.10E+05
PL3 B4	LLGIDVWEHAYYLQYKNVRPDYLKA		1.80E+05
PL3 B5	KNVRPDYLKAIWNVINWENVTERYM		3.60E+05
PL3 B6	ALSRSWQEARLQGVRFLSSREVDRM	**gi15082295**	3.45E+05
PL3 B7	TKKHLNSKTVGQCLETTAQRVPERE		6.30E+05
PL3 B8	PEREALVVLHEDVRLTFAQLKEEVD		1.50E+05
PL3 B9	LVFPKQFKTQQYYNVLKQICPEVEN		2.25E+05
PL3 B10	LKQICPEVENAQPGALKSQRLPDLT		1.80E+05
PL3 B11	GSTRQHLDQLQYNQQFLSCHDPINI		1.50E+05
PL3 B12	SHYNIVNNSNILGERLKLHEKTPEQ		7.50E+04
PL3 C1	IFNGKKALEAISRERGTFLYGTPTM		9.90E+05
PL3 C2	YGTTENSPVTFAHFPEDTVEQKAES		7.50E+04
PL3 C3	YWGEPQKTEEAVDQDKWYWTGDVAT		2.25E+05
PL3 C4	VKDDRMGEEICACIRLKDGEETTVE		1.80E+05
PL3 C5	VFVTNYPLTISGKIQKFKLREQMER		1.59E+06
PL3 C6	ADPRDKALQDYRKKLLEHKEIDGRL	**gi13529265**	3.33E+06
PL3 C7	IDGRLKELREQLKELTKQYEKSEND		1.65E+05
PL3 C8	KSENDLKALQSVGQIVGEVLKQLTE		2.10E+05
PL3 C9	KATNGPRYVVGCRRQLDKSKLKPGT		4.50E+06
PL3 C10	RYLPREVDPLVYNMSHEDPGNVSYS		3.15E+05
PL3 C11	SIVDKYIGESARLIREMFNYARDHQ		4.50E+04
PL3 C12	RRFSEGTSADREIQRTLMELLNQMD		6.15E+05
PL3 D1	TLHRVKMIMATNRPDTLDPALLRPG		6.15E+05
PL3 D2	TLDPALLRPGRLDRKIHIDLPNEQA		5.55E+05
PL3 D3	FVVQEDFMKAVRKVADSKKLESKLD		6.60E+05
PL3 D4	DSDDEMVEEAVEGHLDDDGLPHGFC	**gi18448714**	7.50E+04
PL3 D5	VTYSSTDRFEGNFVHGEKNGRGKFF		5.70E+05
PL3 D6	VYTYEDGGVLQGTYVDGELNGPAQE		7.50E+04
PL3 D7	ELNGPAQEYDTDGRLIFKGQYKDNI		1.35E+05
PL3 D8	EVNEDGEMTGEKIAYVYPDERTALY		1.50E+04
PL3 D9	MSTEEGRPHFELMPGNSVYHFDKST		4.50E+04
PL3 D10	YHFDKSTSSCISTNALLPDPYESER		1.20E+05
PL3 D11	RITHQEVDSRDWALNGNTLSLDEET		3.00E+04
PL3 D12	DVPEPYNHVSKYCASLGHKANHSFT		1.05E+05
PL3 E1	AYGYDHSPPGKSGPEAPEWYQVELK		1.50E+05
PL3 E2	IMDKKANIRNMSVIAHVDHGKSTLT	**gi31107**	7.50E+05
PL3 E3	ARAGETRFTDTRKDEQERCITIKST		2.85E+05
PL3 E4	KSTAISLFYELSENDLNFIKQSKDG		9.00E+04
PL3 E5	MNKMDRALLELQLEPEELYQTFQRI		9.00E+04
PL3 E6	YVAKFAAKGEGQLGPAERAKKVEDM		4.80E+05
PL3 E7	AERAKKVEDMMKKLWGDRYFDPANG		3.00E+05
PL3 E8	GDRYFDPANGKFSKSATSPEGKKLP		7.20E+05
PL3 E9	MNFKKEETAKLIEKLDIKLDSEDKD		1.50E+05
PL3 E10	DIKLDSEDKDKEGKPLLKAVMRRWL		1.50E+05
PL3 E11	HLPSPVTAQKYRCELLYEGPPDDEA		1.50E+05
PL3 E12	SCDPKGPLMMYISKMVPTSDKGRFY		2.10E+05
PL3 F1	GLKVRIMGPNYTPGKKEDLYLKPIQ		6.00E+05
PL3 F2	KDLEEDHACIPIKKSDPVVSYRETV		1.65E+05
PL3 F3	SKSPNKHNRLYMKARPFPDGLAEDI		3.14E+06
PL3 F4	DIDKGEVSARQELKQRARYLAEKYE		9.90E+05
PL3 F5	VLNRKRGHVFEESQVAGTPMFVVKA		1.91E+06

Plate position	Peptide sequence	Gi code	Amounts of eluted phages
PL3 F6	GDPFDNSSRPSQVVAETRKRKGLKE		2.67E+06
PL3 F7	PPVGRRLPGSEPRPELRSFSSEEVI	**gi15866713**	1.01E+06
PL3 F8	PNPSWNKDLRLLFDQFMKKCEDGSW		1.50E+05
PL3 F9	KCEDGSWKRLPSYKRTPTEWIQDFK		1.01E+06
PL3 F10	TPTEWIQDFKTHFLDPKLMKEEQMS		4.50E+04
PL3 F11	AQLFTRSFDDGLGFEYVMFYNDIEK		9.00E+04
PL3 F12	NINYKRPIPLCSVVMINSQLDKVEG		7.50E+04
PL3 G1	KVEGRKFFVSCNVQSVDEKTLYSEA		2.55E+05
PL3 G2	RTNVKKYRYQDEDAPHDHSLPRLTH	**gi1463025**	1.85E+06
PL3 G3	SHSKAVEALKEAGSIARLYVRRRRP		6.75E+05
PL3 G4	GRYSPIPKHMLVDDDYTRPPEPVYS		1.35E+05
PL3 G5	YTRPPEPVYSTVNKLCDKPASPRHY		1.80E+05
PL3 G6	HSQHSTATRQPSMTLQRAVSLEGEP		2.25E+05
PL3 G7	AQYQPEDYARFEAKIHDLREQMMNH		9.00E+04
PL3 G8	SLRTNQKRSLYVRAMFDYDKSKDSG		7.00E+06
PL3 G9	DSEEMGVIPSKRRVERKERARLKTV		9.00E+05
PL3 G10	SKGSFNDKRKKSFIFSRKFPFYKNK		2.34E+06
PL3 G11	SRKFPFYKNKEQSEQETSDPERGQE		6.75E+05
PL3 G12	VPHTTRPKRDYEVDGRDYHFVISRE		9.60E+05
PL3 H1	RDYHFVISREQMEKDIQEHKFIEAG		9.00E+04
PL3 H2	PRSLESLMEMNKRLTEEQAKKTYDR		1.50E+05
PL3 H3	HTDEEIAKHKGPPVFTQEERYKMVQ	**gi12653166**	1.65E+05
PL3 H4	DGRDTYEEVKQAGRYRECKRTQGVS		8.25E+05
PL3 H5	KAHHSSQEMSSEYREYADSFGKCPG		1.50E+05
PL3 H6	TGVSQFLQTSQKIIQFASGKEPQPG		1.05E+05
PL3 H7	LHFDQEVNHYKGKNYPIMNLHERTL		3.30E+05
PL3 H8	IPDRDGSDPYQEPKRRGIFRQIDSG		2.96E+06
PL3 H9	NRLEYEARNQKKEAKELAFLEAARQ		1.49E+06
PL3 H10	AKERRRAVLELLQRPGNARCADCGA	**gi3281993**	2.45E+06
PL3 H11	FYYRPTPSDCQLLREQWIRAKYERQ		2.40E+05
PL3 H12	IYPEKQEPYSAGYREGFLWKRGRDN		1.80E+05
PL4 A1	LSRKFVLTEREGALKYFNRNDAKEP		7.20E+05
PL4 A2	HGLQVTYLKDNSTRNIFIYHEDGKE		4.50E+04
PL4 A3	VPKLSRNYLKEGYMEKTGPKQTEGF		6.45E+05
PL4 A4	KTGPKQTEGFRKRWFTMDDRRLMYF		2.40E+06
PL4 A5	TPDRKFLFACETESDQREWVAAFQK		9.00E+04
PL4 A6	GGPQKGTRFRPLSFEVPKPLFPVAG	**gi13938606**	3.47E+06
PL4 A7	FYQPDEPLTQFLEAAQQEFNLPVRY		7.50E+04
PL4 A8	MLEAHRRQRHPFLLLGTTANRTQSL		1.95E+05
PL4 A9	QSLNYGCIVENPQTHEVLHYVEKPS		1.20E+05
PL4 A10	FSPEALKPLRDVFQRNQQDGQLEDS		7.50E+04
PL4 A11	LYASRLYLSRYQDTHPERLAKHTPG		1.65E+05
PL4 A12	VEGTPSDPNPNDPRARMDSESLFKD		1.05E+05
PL4 B1	RIPAEVLILNSIVLPHKELSRSFTN		2.70E+05
PL4 B2	QKGLKDNFADVQVSVVDCPDLTKEP	**gi12052735**	1.35E+05
PL4 B3	PYLLPLVNQKKVYDLNKIAKEIKLP		7.80E+05
PL4 B4	PVIQTESEHKPPVNGSYFAHVNPAD		1.05E+05
PL4 B5	LEKYSEKCHDFQCALLANLFASEGQ		1.20E+05
PL4 B6	EVKAKRRTGPLNFVTCMRETLEKHY		2.34E+06
PL4 B7	ETLEKHYGNKPIGMGGTFIIQKGKV		3.90E+05
PL4 B8	KGKVKSHIMPAEFSSCPLNSDEEVN		5.25E+05
PL4 B9	RLEHTHFFSRHGEGGHYHYDTTPDI		1.95E+05
PL4 B10	IVEYLGYFLPAEFLYRIDQPKETHS		6.00E+04

Plate position	Peptide sequence	Gi code	Amounts of eluted phages
PL4 B11	AAENEEDEHSVNYKPPAQKSIQEIQ	**gi13543435**	2.40E+05
PL4 B12	PAQKSIQEIQELDKDDESLRKYKEA		4.50E+04
PL4 C1	HTYRKGVKIDKTDYMVGSYGPRAEE		9.30E+05
PL4 C2	SIKSRFTDDDKTDHLSWEWNLTIKK		2.70E+05
PL4 C3	GREKELSIHFVPGSCRLVEEEVNIP	**gi12052851**	6.00E+04
PL4 C4	VHKEFQQNNWHAVGCGFRRARPKFE		1.70E+06
PL4 C5	AAERRPDVVENQPDAASQLNVDASG		7.50E+04
PL4 C6	DGTNPPYREEDIPAPLNLYGKTKLD		7.50E+04
PL4 C7	GEVEKLEESAVTVMFDKVQFSNKSA		9.00E+04
PL4 C8	DKVQFSNKSANMDHWQQRFPTHVKD		1.80E+05
PL4 C9	FPTHVKDVATVCRQLAEKRMLDPSI		1.65E+05
PL4 C10	AEKRMLDPSIKGTFHWSGNEQMTKY		2.25E+05
PL4 C11	GAQRPRNAQLDCSKLETLGIGQRTP		6.30E+05
PL4 C12	GQRTPFRIGIKESLWPFLIDKRWRQ		3.00E+05
PL4 D1	GPMRKSYRGDREAFEETHLTSLDPV	**gi13676377**	1.05E+06
PL4 D2	CTRDGKPSARMLLLKGFGKDGFRFF		1.01E+07
PL4 D3	NFESRKGKELDSNPFASLVFYWEPL		1.65E+05
PL4 D4	VRVEGPVKKLPEEEAECYFHSRPKS		1.05E+05
PL4 D5	DREYLRKKNEELEQLYQDQEVPKPK		7.50E+04
PL4 D6	RRGLPTGDSPLGPMTHRGEEDWLYE		6.45E+05
PL4 D7	QVSSKVKSSPTWYSASSFSSSVPTV	**gi7160984**	2.55E+05
PL4 D8	GKFVESKSDKWIDIHNPATNEVIGR		1.80E+05
PL4 D9	NPATNEVIGRVPQATKAEMDAAIAS		6.00E+04
PL4 D10	RYQQLIKENLKEIAKLITLEQGKTL		9.00E+05
PL4 D11	LMKPSERVPGATMLLAKLLQDSGAP		1.65E+05
PL4 D12	SNKAGEYIFERGSRHGKRVQANMGA		2.94E+06
PL4 E1	KRVQANMGAKNHGVVMPDANKENTL		2.42E+06
PL4 E2	DLGPLITPQAKERVCNLIDSGTKEG		2.25E+05
PL4 E3	DSGTKEGASILLDGRKIKVKGYENG		9.75E+05
PL4 E4	LETETLDEAIQIVNNNPYGNGTAIF		4.50E+04
PL4 E5	NPYGNGTAIFTTNGATARKYAHLVD		1.05E+05
PL4 E6	FYGKQGIQFYTQLKTITSQWKEEDA		7.50E+04
PL4 E7	AAEEEPKPKKLKVEAPQALRENILF	**gi1906008**	2.55E+05
PL4 E8	KDFLDKYSLKPNDQILAEDKHKELF		1.05E+05
PL4 E9	QPHKAATFFGCIGIDKFGEILKRKA		3.30E+05
PL4 E10	FGEILKRKAAEAHVDAHYYEQNEQP		2.85E+05
PL4 E11	GDNRSLIANLAAANCYKKEKHLDLE		3.75E+05
PL4 E12	HASENNRIFTLNLSAPFISQFYKES		5.40E+05
PL4 F1	FAREQGFETKDIKEIAKKTQALPKM		6.60E+05
PL4 F2	AKKTQALPKMNSKRQRIVIFTQGRD		3.89E+06
PL4 F3	QGRDDTIMATESEVTAFAVLDQDQK		1.95E+05
PL4 F4	TECIRAGHYAASIIIRRTGCTFPEK		6.00E+04
PL4 F5	GFAGDQIPKYCFPNYVGRPKHMRV	**gi563885**	2.40E+05
PL4 F6	PKAEEHRGLLTIRYPMEHGVVRDWN		1.50E+05
PL4 F7	RDWNDMERIWQYVYSKDQLQTFSEE		7.50E+04
PL4 F8	QTFSEEHPVLLTEAPLNPSKNREKA		1.95E+05
PL4 F9	VVRTIKERACYLSINPQKDEALETE		1.50E+05
PL4 F10	PQKDEALETEKVQYTLPDGSTLDVG		4.50E+04
PL4 F11	LLSEVKKLAPKDIKIKISAPQERLY		4.80E+05
PL4 F12	ASLDTFKKMWVSKKEYEEDGSRAIH		3.00E+05
PL4 G1	LEDDEDMTLTRWTGMIIGPPRTIYE	**gi12653396**	1.20E+05
PL4 G2	PPRTIYENRIYSLKIECGPKYPEAP		1.20E+05
PL4 G3	AISVLAKWQNSYSIKVVLQELRRLM		1.05E+05

Plate position	Peptide sequence	Gi code	Amounts of eluted phages
PL4 G4	VVLQELRRLMMSKENMKLPQPPEGQ		7.50E+04
PL4 G5	SDSEDSNFSEEEDSERSSDGEEAEV	**gi2065176**	1.50E+05
PL4 G6	RSSDGEEAEVDEERRSAAGSEKEEE		6.00E+04
PL4 G7	SAAGSEKEEEPEDEEEEEEEEEYDE		1.05E+05
PL4 G8	EEEEEEEYDEEEEEEDDDRPPKKPR		1.20E+05
PL4 G9	KKPRHGGFILDEADVDDEYEDEDQW		4.50E+05
PL4 G10	ARRLQNLWRDQREEELGEYYMKKYA		4.80E+05
PL4 G11	DRIKARMSLKDWFAKRKKFKRPPQR		4.44E+06
PL4 G12	KIRSLGGDVASDGDFLIFEGNRYSR		2.40E+05
PL4 H1	GVKPTLSELEKFEDQPEGIDLEVVT	**gi2065176**	1.05E+05
PL4 H2	PEGIDLEVVTESTGKEREHNFQPGD		1.20E+05
PL4 H3	RHQAVTRKKDNRFAVALDSEQNNIH		3.00E+06
PL4 H4	SRGRGRRDNELIGQTVRISQGPYKG		4.44E+06
PL4 H5	AWDPNNPNTPSRAEEEYEYAFDDEP		2.25E+05
PL4 H6	EYEYAFDDEPTPSPQAYGGTPNPQT		2.10E+05
PL4 H7	FDDEPTPSPQAYGGTPNPQTPGYPD		7.50E+04
PL4 H8	PNPQTPGYPDPSSPQVNPQYNPQTP		1.80E+05
PL4 H9	LKDSEKVVSISSEHLEPITPTKNNK		1.50E+05
PL4 H10	KSLKKLVEESREKNQPEVDMSDRGI	**gi14250498**	6.30E+05
PL4 H11	LPPELGNLDLTGQKQVFKAENNPWV		1.20E+05
PL4 H12	FEYIRSETYKYLYGRHMQANPEPPK		6.00E+04
PL5 A1	HMQANPEPPKKNNDKSKKISRKPLA		2.04E+06
PL5 A2	STRTLGGRSLSTAERRAEPGARRPR	**gi20521841**	3.23E+06
PL5 A3	AFPPRAKDGLVVFGKNSARPRDEVQ		1.11E+06
PL5 A4	NSARPRDEVQEVVYFSAADHEPESK		1.05E+05
PL5 A5	AEMGANEHGVCIANEAINTREPAAE		9.00E+04
PL5 A6	GKDSLEKQEESITVQTMMNTLRDKA		1.20E+05
PL5 A7	VLPQNRSSPCIHYFTGTPDPSRSIF		1.05E+05
PL5 A8	KLVPKTQSPCFGDDDPAKKEPRFQE		1.80E+05
PL5 A9	PAKKEPRFQEKPDRRHELYKAHEWA		1.22E+06
PL5 A10	HELYKAHEWARAIIESDQEQGRKLR		6.00E+04
PL5 A11	VENEEASASIIVKMTDSFTEQADQV	**gi12653600**	6.00E+04
PL5 A12	AGGNAKSKSLFKNCDLMWKQSIWTS		3.11E+06
PL5 B1	LMWKQSIWTSTISSHLATKHLKEGG		3.30E+05
PL5 B2	TLDTPMNRKSMPEADFSSWTPLEFL		2.70E+05
PL5 B3	ITGKNRPSSGSLIQVVTTEGRTELT		5.70E+05
PL5 B4	LGEAYATKARAHGLEVEPSALEQGF	**gi13477172**	3.75E+05
PL5 B5	GTPPPPYTVG	control	2.19E+06
PL5 B6	GTPPPPYTVG		1.80E+06
PL5 B7	GTPPPPYTVG		1.31E+06
PL5 B8	GTPPPPYTVG		1.26E+06

Appendix 4

Sequences of proteins identified using the microarray approach.

Nucleotide sequence of identified clone: DKFZp761O0120Q:

Library 761, Plate 20, row O, column 1

5′ TGAATTTAGGTGACACTATAGAAGAGCTATGACGTCGCATGCACGCGTACGTAAGCTTGGATCCTCTAGAGCG
GCCGCTTTTTTTTTTTTTTTTTTTCAACTATTGTATTTATTTAAAACAAACTGACAAGTTCAAGCACCTGTCTTC
AGAAAAGCCAGCAGCATTTTTTTTTTTTTAACATACTCAAAGTAAGATTTGGCCTAAGCCCTTAATACCTTTCTG
AACAGCCATGCAACTAAACACCCTCAGGAGATGTTACATAAGGGAGAGAAGAACATGGAGCAATTTGCACTTTTT
CCCCTAGATAATATTAACAAGGTAAAGCAAATCCAGATCTTTATGAATGAATGGCTGTCATGTTTAATACACTTG
GAGCTCTATAAAACTAGAGCCACTATCATATATGTTTATATAGATATATATGTATTTTTAATAAAGTCTCTCACT
TGCTTCAAAATTTAAGGATGATCTTTTTCCTGGTCAGTCATTTCAAGGGAACAATAAATAAAGTAGAAAAGGATG
AACTGTTTTGCATATTTCTGTGGAAGAACGGAAAACATACTTTGAGAGGTTCGGAAAATTCACTGTACAGAATGA
ACAGCTTATATACACTATCTGTGCAATTTTTCTTGGCTGGGGTATTTAAATGCAAGAGCTCCTCCAAATTAAGAA
GCCACATTCATTCTTGTCCAAATCCCTCTGTCTCTTCTATTGCAAAAAGAAGTTTTCCTTTAGTTGTTCATAACT
CTTAATNGGNGGTAGATCCAAGCGATAAAACNTGTATGGCTTCTTGGGNANCNANTGTCTTTGCCAACTTTTTNA
ATGCAAACTTTTGAGGCCCNTTCTTCCNAGAGCTAGCAATCCTCCTAAGGTAACGGCGGTCCAGGACNAATGCAA
AG **3′**

Protein sequences: Nedd-4-like E3 ubiquitin-protein ligase, identified as interaction partner for peptide ligand originating from NEDD4-like ubiquitin ligase, using microarray approach 1. **WW1**, **WW2**, **WW3**, **WW4**, **QP** – overlap of two domains, **RLDL*PPYP*SYSMLYEKLLTAVEETS** peptide 107.

>sp|Q9H0M0|WWP1_HUMAN Nedd-4-like E3 ubiquitin-protein ligase WWP1 (EC 6.3.2.-) (WW domain- containing protein 1) (Atropin-1 interacting protein 5) (AIP5) - Homo sapiens (Human).

MATASPRSDTSNNHSGRLQLQVTVSSAKLKRKKNWFGTAIYTEVVVDGEITKTAKSSSSS
NPKWDEQLTVNVTPQTTLEFQVWSHRTLKADALLGKATIDLKQALLIHNRKLERVKEQLK
LSLENKNGIAQTGELTVVLDGLVIEQENITNCSSSPTIEIQENGDALHENGEPSARTTAR
LAVEGTNGIDNHVPTSTLVQNSCCSYVVNGDNTPSSPSQVAARPKNTPAPKPLASEPADD
TVNGESSSFAPTDNASVTGTPVVSEENALSPNCTSTTVEDPPVQEILTSSENNECIPSTS
AELESEARSILEPDTSNSRSSSAFEAAKSRQPDGCMDPVRQQSGNANT**ETLPSGWEQRKD**
PHGRTYYVDHNTRTTTWERPQPLPPGWERRVDDRRRVYYVDHNTRTTTWQRPTMESVRNF
EQWQSQRNQLQGAMQQFNQRYLYSASMLAAENDPY**GPLPPGWEKRVDSTDRVYFVNHNTK**
TTQWEDPRTQGLQNE**EPLPEGWEIRYTREGVRYFVDHNTRTTTFKDPRN**GKSSVTKGGPQ
IAYERGFRWKLAHFRYLCQSNALPSHVKINVSRQTLFEDSFQQIMALKPYDLRRRLYVIF
RGEEGLDYGGLAREWFFLLSHEVLNPMYCLFEYAGKNNYCLQINPASTINPDHLSYFCFI
GRFIAMALFHGKFIDTGFSLPFYKRMLSKKLTIKDLESIDTEFYNSLIWIRDNNIEECGL
EMYFSVDMEILGKVTSHDLKLGGSNILVTEENKDEYIGLMTEWRFSRGVQEQTKAFLDGF
NEVVPLQWLQYFDEKELEVMLCGMQEVDLADWQRNTVYRHYTRNSKQIIWFWQFVKETDN
EVRMRLLQFVTGTCRLPLGGFAELMGSNGPQKFCIEKVGKDTWLPRSHTCFNRLDLPPYK
SYEQLKEKLLFAIEETEGFGQE

>tr|Q9HCC7|Q9HCC7_HUMAN NEDD4-like ubiquitin ligase 1 - Homo sapiens (Human).

MASPSRNSQSRRRCKEPLRYSYNPDQFHNMDLRGGPHDGVTIPRSTSDTDLVTSDSRSTL
MVSSSYYSIGHSQDLVIHWDIKEEVDAGDWIGMYLIDEVLSENFLDYKNRGVNGSHRGQI
IWKIDASSYFVEPETKICFKYYHGVSGALRATTPSVTVKNSAAPIFKSIGADETVQGQGS
RRLISFSLSDFQAMGLKKGMFFNPDPYLKISIQPGKHSIFPALPHHGQERRSKIIGNTVN
PIWQAEQFSFVSLPTDVLEIEVKDKFAKSRPIIKRFLGKLSMPVQRLLERHAIGDRVVSY

TLGRRLPTDHVSGQLQFRFEITSSIHPDDEEISLSTEPESAQIQDSPMNNLMESGSGEPR
SEAPESSESWKPEQLGEGSVPDRPGNQSIELSRPAEEAAVITEAGDQGMVSVGPEGAGEL
LAQVQKDIQPAPSAEELAEQLDLGEEASALLLEDGEAPASTKEEPLEEEATTQSRAGREE
EEKEQEEEGDVSTLEQGEGRLQLRASVKRKSRPCSLPVSELETVIASACGDPETPRTHYI
RIHTLLHSMPSAQGGSAAEEEDGAEEESTLKDSSEKDGLSEVDTVAADPSALEEDREEPE
GATPGTAHPGHSGGHFPSLANGAAQDGDTHPSTGSESDSSPRQGGDHSCEGCDASCCSPS
CYSSSCYSTSCYSSSCYSASCYSPSCYNGNRFASHTRFSSVDSAKISESTVFSSQDDEEE
ENSAFESVPDSMQSPELDPESTNGAGPWQDELAAPSGHVERSPEGLESPVAGPSNRREGE
CPILHNSQPVSQLPSLRPEHHHYPTID**EPLPPNWEARIDSHGRVFYVDHVNRTTTWQRPT**
AAATPDGMRRSGSIQQMEQLNRRYQNIQRTIATERSEEDSGSQSCEQAPAGGGGGGGSDS
EAESSQSSLDLRREGSLSPVNSQKITLLLQSPAVKFITNPEFFTVLHANYSAYRVFTSST
CLKHMILKVRRDARNFERYQHNRDLVNFINMFADTR**LELPRGWEIKTDQQGKSFFVDH**NS
RATTFIDPRIPLQNGRLPNHLTHRQHLQRLRSYSAGEASEVSRNRGASLLARPGHSLVAA
IRSQHQHESLPLAYNDKIVAFLRQPNIFEMLQERQPSLARNHTLREKIHYIRTEGNHGLE
KLSCDADLVILLSLFEEEIMSYVPLQAAFHPGYSFSPRCSPCSSPQNSPGLQRASARAPS
PYRRDFEAKLRNFYRKLEAKGFGQGPGKIKLIIRRDHLLEGTFNQVMAYSRKELQRNKLY
VTFVGEEGLDYSGPSREFFFLLSQELFNPYYGLFEYSANDTYTVQISPMSAFVENHLEWF
RFSGRILGLALIHQYLLDAFFTRPFYKALLRLPCDLSDLEYLDEEFHQSLQWMKDNNITD
ILDLTFTVNEEVFGQVTERELKSGGANTQVTEKNKKEYIERMVKWRVERGVVQQTEALVR
GFYEVVDSRLVSVFDARELELVIAGTAEIDLNDWRNNTEYRGGYHDGHLVIRWFWAAVER
FNNEQRLRLLQFVTGTSSVPYEGFAALRGSNGLRRFCIEKWGKITSLPRAHTCFN**RLDLP**
PYPSYSMLYEKLLTAVEETSTFGLE

Nucleotide sequence of identified clone: DKFZp761L0517Q. Library 761

Library 761, Plate 17, row L, column 5

GAATTGAATTTAGGTGACACTATAGAAGAGCTATGACGTCGCATGCACGCGTACGTAAGCTTGGATCCTCTAGAG
CGGCCGCATCCTGGCAGCAGCGGGACAGCGTCTCAGCTCCCCCACCTCCCAGTCCCCTAGCACCTTTCTAGCTCA
TCCGGCTGTGTGGCCTTGGGCAAGGGCAGGAACCTCTCTGAGCTTCAGTTTCCTCATCTACGCAATGGGGCTGTG
CGAGAACACAAGCCGTTCATGGAGATGCCACTCCTCAGGTGGAGGCTGCCAGCTCTGGGCTTTGGTGCCAGCTGT
CATTTATGGAAGAGCAGGCCGCCCTGCCCTGGCCTCATCAGCCCCAGCAGCCCCTCTGGCTTCCTCATGGGCCGT
GTGACCGCAGGGGCTTCCTGACACACATGGGTGCGAGTGTCCCAGGACCAGCCTCTACTTGCTACTGTGCTGTCT
TCTCGAGTGGGAACTCCACTGCAAACATGGGGAGGAAGGGAAGGAGCTCGCTGTGGGAGTCTGGGCGCGGTGGAC
GTTAACAATGGGGGGCGAAGTGTCCTCCCACGGTGGGTGCACTGTCTGTGGGAAGGCTGACCACGCTGTTAGCTT
TTCAGGGATCTTGGTTGGCCTCCTGAGTCTCTACTCAATGGACAGGGTCCAGACTGCACTTTACATTTTGCCAAC
AAGAAATGGGAGACTGTCCTTGAGCCCTCTCCTGAAGACTTGAGGGAACCAAACCAGACTGTCCCTCCTCTATCA
GCCTGCAGCCTTTCTTCCATTTCATTAAAAGTGTTTAGGACAGGGAGGACCTTGACAACCTAGCCCATCTCTTAT
TAGAGGGGAGGAGATGACCCAGAGGCCTCAAGcTTcCAGAGcaCTGGNTGCACCCNAGCCTCTGCTGGGGGTTCC

Myelin Basic Protein: positively enriched on peptide-ligand for WW/FE65

>CAH10359, NM_001025081 Myelin basic protein (MBP)
MASQKRPSQRHGSKYLATASTMDHARHGSLPQKSHGRTQDENPVVHFFKNIVTPRTPPPS
QGKGRGLSLSRFSW

Appendix 5

Lists of genes selected from separate microarray experiments using different values of post-/pre-panning ratio. Filtration thresholds: signal ≥ 0.1; CV ≤ 0.5; ratio ≥ 15; ratio ≥ 10; ratio ≥ 5. ***Bold, italic, underlined***: genes present only in "sister" experiments or showing high. **Bold underlined**: genes present in "sister" and other experiments.

Pannings on pep-E

WWYAP suppl 1x 10^5	Post-/pre-panning ratio	WWYAP suppl 1x 10^4	Post-/pre panning ratio	WWYAP no suppl	Post-/pre-panning ratio
DKFZp761A212Q	11.41414141	DKFZp761D1915Q	6.420212766	DKFZp761A092Q	5.14489
DKFZp761B0313Q	11.04093567	DKFZp761H059Q	5.828877005	DKFZp761A098Q	11.48913
DKFZp761B1915Q	13.25	DKFZp761I112Q	5.572052402	DKFZp761A099Q	5.474026
DKFZp761C099Q	10.69444444	DKFZp761I199Q	5.844919786	DKFZp761A138Q	6.722054
DKFZp761C1516Q	16.25396825	DKFZp761M037Q	7.198312236	DKFZp761A218Q	9.396947
DKFZp761D0712Q	12.68421053	DKFZp761M199Q	7.552325581	DKFZp761A219Q	5.453165
DKFZp761D0713Q	14.65822785	DKFZp761M217Q	9.847457627	DKFZp761D138Q	6.85567
DKFZp761D0716Q	13.3442623	DKFZp761N0815Q	5.037593985	DKFZp761E0713Q	7.088968
DKFZp761D0812Q	10.04807692	***DKFZp761O0120Q***	***72.875***	DKFZp761E1716Q	10.6242
DKFZp761D0816Q	10.38095238	DKFZp761P1610Q	5.240339303	***DKFZp761F1212Q***	***47.42951***
DKFZp761D2310Q	21.20754717	E	21.4173913	DKFZp761F1510Q	161.2387
DKFZp761E1612Q	12.03030303	Y	31.80769231	DKFZp761F1611Q	44.59633
DKFZp761F1212Q (1)	***10.29180328***			DKFZp761F187Q	6.040609
DKFZp761F1913Q	14.62790698			DKFZp761F2010Q	5.428439
DKFZp761G037Q	11.64705882			DKFZp761G108Q	5.340206
DKFZp761G1715Q	12.75586854			DKFZp761H0617Q	11.65019
DKFZp761G1820Q	12.8220339			DKFZp761H2418Q	6.475169
DKFZp761G247Q	13.13541667			DKFZp761I0215Q	6.812155
DKFZp761H2319Q	12.36842105			DKFZp761I168Q	5.486111
DKFZp761I0213Q	10.25			DKFZp761I1715Q	7.435294
DKFZp761I077Q	10.5			DKFZp761I187Q	6.539535
DKFZp761J068Q	13.06329114			DKFZp761I2115Q	7.355401
DKFZp761J1212Q	13.44303797			DKFZp761J0210Q	6.991304
DKFZp761K0118Q	14.12658228			DKFZp761J0313Q	5.501319
DKFZp761K032Q	11.59615385			DKFZp761J068Q	13.79747
DKFZp761K048Q	10.04237288			DKFZp761K0217Q	6.528139
DKFZp761K1117Q	10.09195402			DKFZp761L1019Q	5.797753
DKFZp761K2310Q	11.88636364			DKFZp761M018Q	5.21659
DKFZp761K231Q	10.20769231			**DKFZp761M037Q**	**5.472574**
DKFZp761L2112Q	19.09259259			DKFZp761M088Q	5.212544
DKFZp761L242Q	25.16901408			DKFZp761M0912Q	5.631297
DKFZp761N1116Q	12.53488372			DKFZp761M097Q	7.862205
DKFZp761N2013Q	12.7826087			DKFZp761M118Q	5.165217
DKFZp761O0120Q (2)	***81.375***			DKFZp761M1713Q	5.943231
DKFZp761O0216Q	18.5			DKFZp761M172Q	10.5
DKFZp761O2113Q	18.23880597			DKFZp761M2015Q	6.169192
DKFZp761O2412Q	19.49090909			DKFZp761M217Q	10.36441
DKFZp761P012Q	10.00724638			DKFZp761M239Q	6.847134
DKFZp761P0617Q	11.54421769			***DKFZp761O0120Q***	***99.0625***
DKFZp761P1413Q	10.95959596			DKFZp761O0614Q	5.608541
DKFZp761P1811Q	14.42857143			DKFZp761O152Q	5.107527

WWYAP suppl 1x 10^5	Post-/pre-panning ratio	WWYAP suppl 1x 10^4	Post-/pre panning ratio	WWYAP no suppl	Post-/pre-panning ratio
DKFZp761P2312Q	11.04807692			DKFZp761O1715Q	7.795455
E	34.8				
Y	26.85576923				

(1) Human DNA sequence from clone RP11-127L20 on chromosome 10 Contains the gene for a novel glutathione-S-transferase gene(LOC119391), a novel gene (KIAA1754), the 5'end of a novel gene (FLJ35908), two novel genes and two CpG islands, complete sequence.

(2) sequenced gene present also in experiment 107B. Q9HCC7|Q9HCC7_HUMAN NEDD4-like ubiquitin ligase 1

Pannings on pep-F

Pan WW/FE65 suppl 1x 10^5 (exp 1)	Post-/pre-panning ratio	Pan WW/FE65 suppl 1x 10^4 (exp 2)	Post-/pre-panning ratio
DKFZp761C1516Q	18.12698413	DKFZp761B108Q	5.913978495
DKFZp761C2012Q	23.95238095	DKFZp761D131Q	5.644657863
DKFZp761C208Q	16.09090909	DKFZp761E0217Q	5.062602965
DKFZp761D0519Q	26.03448276	DKFZp761G088Q	5.184649611
DKFZp761D1314Q	19.07936508	DKFZp761H0513Q	8.720670391
DKFZp761D1418Q	17.34482759	DKFZp761H0614Q	5.10617284
DKFZp761D1620Q	31.36111111	DKFZp761H2418Q	5.888261851
DKFZp761D1820Q	18.45	DKFZp761I058Q	5.401337793
DKFZp761D2310Q	26.71698113	DKFZp761J152Q	5.193675889
DKFZp761H0516Q	18.66129032	***DKFZp761L0517Q***	***15.47945205***
DKFZp761H0814Q	36.18918919	DKFZp761L1810Q	5.923728814
DKFZp761H0819Q	16.26470588	DKFZp761M1316Q	5.515454545
DKFZp761H1319Q	23.20454545	DKFZp761M199Q	5.837209302
DKFZp761H1510Q	19.23076923	DKFZp761M2017Q	5.141025641
DKFZp761H1618Q	22.18461538	DKFZp761P0311Q	5.875
DKFZp761H2119Q	17.33333333	DKFZp761P1115Q	9.923809524
DKFZp761H2319Q	17.11578947	DKFZp761P1510Q	7.77443609
DKFZp761H2420Q	22.90909091	DKFZp761P1610Q	6.058435438
DKFZp761K0118Q	18.27848101	DKFZp761P2114Q	10.07142857
DKFZp761K1920Q	25.95	DKFZp761P2119Q	5.09569378
DKFZp761L0517Q (3)	***15.56164384***	E	24.63478261
DKFZp761L242Q	19.76056338		
DKFZp761O1110Q	17.43103448		
DKFZp761O1715Q	15.25		
DKFZp761P0219Q	19.37096774		
DKFZp761P032Q	20.57142857		
DKFZp761P1019Q	16.3880597		
E	94.24347826		
Ezr	24.85714286		
F	33.62790698		

(3) Not verified gene sequence

Pannings on pep-E

EVH1 suppl. 1x 10^4	Post-/pre-panning ratio	EVH1 no suppl.	Post-/pre-panning ratio
DKFZp761A052Q	10.65263158	DKFZp761A137Q	13.4516129
DKFZp761A0810Q	10.11764706	DKFZp761A218Q	10.22900763
DKFZp761B199Q	13.38666667	DKFZp761C0817Q	12.0703125
DKFZp761C1516Q	16.3968254	DKFZp761D0812Q	10.66346154

EVH1 suppl. 1x 10^4	Post-/pre-panning ratio	EVH1 no suppl.	Post-/pre-panning ratio
DKFZp761C247Q	16.96969697	DKFZp761E1716Q	10.13375796
DKFZp761D0716Q	10.99180328	DKFZp761G247Q	12.26041667
DKFZp761D178Q	11.13888889	DKFZp761I077Q	10.78846154
DKFZp761D2310Q	18.88679245	DKFZp761J228Q	10.28703704
DKFZp761E1612Q	11.4469697	DKFZp761M2013Q	10.36184211
DKFZp761F0318Q	15.92307692	**DKFZp761M211Q**	**12.0610687**
DKFZp761F1913Q	14.1627907	**DKFZp761O1715Q**	**12.25**
DKFZp761G247Q	11.44791667		
DKFZp761H2319Q	10.71578947		
DKFZp761K0118Q	18.03797468		
DKFZp761K051Q	11.67045455		
DKFZp761K2310Q	11.69318182		
DKFZp761L027Q	15.39534884		
DKFZp761L2015Q	10.65957447		
DKFZp761M211Q	**10.13740458**		
DKFZp761N1116Q	14.29069767		
DKFZp761O012Q	12.65420561		
DKFZp761O0313Q	11.2125		
DKFZp761O1715Q	**21.32575758**		
E	43.60869565		

Pannings on other peptides, without supplement of model phage

200 A	Post-/pre-panning ratio	200 B
DKFZp761E151Q	8.267310789	0
DKFZp761F081Q	5	
DKFZp761G0716Q	7.476683938	
DKFZp761G1316Q	11.775	
DKFZp761H0513Q	7.391061453	
DKFZp761H0614Q	9.390123457	
DKFZp761K221Q	6.330033003	
DKFZp761M212Q	5.060185185	
DKFZp761O2419Q	5.932153392	
DKFZp761P0311Q	5.462962963	
DKFZp761P2114Q	6.673469388	

199 B	Post-/pre-panning ratio	199 A
DKFZp761E1315Q	5.935672515	hybridization failed
DKFZp761E1716Q	6.636942675	
DKFZp761J0210Q	8.430434783	

148 A	Post-/pre-panning ratio	148 B	Post-/pre-panning ratio
DKFZp761M211Q	7.687022901	**DKFZp761J148Q**	14.01388889

113 A	Post-/pre-panning ratio	113 B	Post-/pre-panning ratio
DKFZp761A0620Q	5.603015075	**DKFZp761A098Q**	**6.14673913**
DKFZp761A098Q	**7.168478261**	DKFZp761A1317Q	6.03196347
DKFZp761B0316Q	5.167420814	DKFZp761B0313Q	6.228070175
DKFZp761B108Q	5.838709677	**DKFZp761B208Q**	**5.904522613**
DKFZp761B152Q	5.459574468	**DKFZp761D138Q**	**6.525773196**
DKFZp761B208Q	**5.472361809**	**DKFZp761E0713Q**	**5.797153025**
DKFZp761B228Q	5.823033708	DKFZp761E1315Q	8.912280702
DKFZp761C119Q	7.069124424	DKFZp761E177Q	5.071428571
DKFZp761D138Q	**6.067010309**	DKFZp761E1915Q	17.80645161
DKFZp761E0713Q	**5.736654804**	**DKFZp761F111Q**	**5.628415301**
DKFZp761E1716Q	6.859872611	DKFZp761F187Q	6.578680203
DKFZp761F027Q	6.974025974	DKFZp761G082Q	5.282945736
DKFZp761F111Q	**5.715846995**	DKFZp761H028Q	5.453441296
DKFZp761F187Q	7.487309645	DKFZp761H059Q	5.663101604
DKFZp761G0315Q	5.707627119	DKFZp761H2020Q	6.704433498
DKFZp761G081Q	5.362573099	DKFZp761I058Q	5.682274247
DKFZp761I0513Q	5.009852217	DKFZp761I099Q	5.064748201
DKFZp761I0617Q	8.090909091	DKFZp761I191Q	5.049180328
DKFZp761I2115Q	**5.212543554**	**DKFZp761I2115Q**	**5.222996516**
DKFZp761J0210Q	**6.904347826**	**DKFZp761J0210Q**	**7.356521739**
DKFZp761J148Q [4]	***26.55555556***	***DKFZp761J148Q***	***17.16666667***
DKFZp761J152Q	5.225296443	**DKFZp761J248Q**	**5.039525692**
DKFZp761J248Q	**5.359683794**	DKFZp761M0316Q	6.040935673
DKFZp761M037Q	6	DKFZp761M047Q	6.333333333
DKFZp761M097Q	**5.598425197**	**DKFZp761M097Q**	**7.275590551**
DKFZp761M2017Q	**5.269230769**	DKFZp761M122Q	5.307692308
DKFZp761N157Q	7.162561576	**DKFZp761M2017Q**	**5.085470085**
DKFZp761N238Q	6.605381166	DKFZp761M201Q	5.221374046
DKFZp761O0112Q	5.335078534	DKFZp761M218Q	5.176470588
DKFZp761O192Q	**20.61538462**	**DKFZp761O192Q**	**14.52564103**
DKFZp761O197Q	5.792079208	DKFZp761P2114Q	7.959183673

[4] NM_018700 Homo sapiens tripartite motif-containing 36 (TRIM36), mRNA

107 B	Post-/pre-panning ratio	107 A
DKFZp761A098Q	7.331521739	Hybridization failed
DKFZp761A218Q	8.015267176	
DKFZp761B108Q	5.555555556	
DKFZp761B152Q	5.69787234	
DKFZp761B208Q	6.052763819	
DKFZp761B228Q	5.261235955	
DKFZp761C048Q	6.147268409	
DKFZp761C0720Q	5.172932331	
DKFZp761C0817Q	9.921875	
DKFZp761C119Q	6.35483871	
DKFZp761C2415Q	5.064935065	
DKFZp761D138Q	6.768041237	
DKFZp761D172Q	5.207964602	
DKFZp761D188Q	5.972055888	
DKFZp761D248Q	6.462908012	
DKFZp761E0713Q	6.174377224	
DKFZp761E1112Q	7.005524862	

107 B	Post-/pre-panning ratio	107 A
DKFZp761E177Q	5.092436975	
DKFZp761F0716Q	5.273255814	
DKFZp761F111Q	6.557377049	
DKFZp761F187Q	9.294416244	
DKFZp761F222Q	6.358490566	
DKFZp761G0715Q	5.255506608	
DKFZp761G081Q	6.204678363	
DKFZp761G082Q	5.360465116	
DKFZp761G1715Q	5.798122066	
DKFZp761G191Q	6.577922078	
DKFZp761G227Q	9.467948718	
DKFZp761H022Q	6.9	
DKFZp761H0513Q	8.268156425	
DKFZp761H088Q	5.113835377	
DKFZp761H118Q	5.107744108	
DKFZp761H142Q	5.333333333	
DKFZp761H169Q	5.171521036	
DKFZp761I0513Q	5.103448276	
DKFZp761I0617Q	7	
DKFZp761I099Q	5.881294964	
DKFZp761I2115Q	5.522648084	
DKFZp761J0210Q	7.3	
DKFZp761J1212Q	6.512658228	
DKFZp761J1413Q	34.09090909	
DKFZp761J148Q	**33.73611111**	
DKFZp761J168Q	5.103969754	
DKFZp761J248Q	5.288537549	
DKFZp761K032Q	6.583333333	
DKFZp761K132Q	5.010869565	
DKFZp761K168Q	5.577287066	
DKFZp761K208Q	7.9375	
DKFZp761L108Q	7.874458874	
DKFZp761L117Q	10.74166667	
DKFZp761L211Q	5.160869565	
DKFZp761M019Q	6.125984252	
DKFZp761M037Q	5.46835443	
DKFZp761M0916Q	5.418719212	
DKFZp761M091Q	5.011415525	
DKFZp761M097Q	5.519685039	
DKFZp761M098Q	5.151436031	
DKFZp761M132Q	8.322834646	
DKFZp761M211Q	8.801526718	
DKFZp761M218Q	6.214285714	
DKFZp761N238Q	6.650224215	
DKFZp761O0120Q	***65.09375***	
DKFZp761O0315Q	5.781818182	
DKFZp761O139Q	10.11111111	
DKFZp761O1715Q	13.57575758	
DKFZp761O192Q	23.29487179	

ACKNOWLEDGEMENTS

Work presented in this thesis was performed in the Department of Chemical Biology in the Chemical Biology (CBIO) group of Dr. Ronald Frank at the German Research Center for Biotechnology (GBF) in Braunschweig. Facilities of many laboratories have been utilized in the process of realization of this PhD program. I would like to thank to all persons who supported me generously in fulfilling my goals, in particular to:

Dr. Ronald Frank, CBIO at the GBF, for giving me the opportunity to work and learn in his laboratory, critical reweaving of this thesis and my other scientific achievements and to entrust this fascinating project to me.

Prof. Dr. Jürgen Wehland, Department of Cell Biology and Immunology (ZIB) of the GBF, for being mentor of this thesis and its reweaving and evaluation.

Prof. Dr. Stefan Dübel, Institute for Biochemistry and Biotechnology of the Technical University of Braunschweig (TU) for reweaving and evaluation of this thesis.

Prof. Dr. Norbert F. Käufer, Institute for Genetics of the Technical University of Braunschweig (TU), for evaluation of my work.

I would like to give my very special thanks to Dr. Klaus-Dieter Aumann, Rechenzentrum at the GBF, for his generous help in various approaches concerned with the bioinformatics, including, high throughput sequence BLAST homology search, large scale downloading of sequences for ASS analysis, for general computational support, for being a good friend and his endless patience, printing of this thesis and support in administrative work necessary for its submission.

Great thanks to Lesley Mühle (TA), CBIO at the GBF, for her continuous technical assistance in the laboratory work, for being very friendly and great patience in teaching me the German language.

Great thanks to Krzysztof Białek (MSc), CBIO at the GBF, for close cooperation in the frame of this project, fruitful discussion and problem troubleshooting.

Thanks to Dr. Antonius Dikmans, CBIO at the GBF, for his extraordinary sense of humor, help with troubleshooting related to robotic systems and patient reweaving of this thesis.

Dr. René Rübenhagen, CBIO at the GBF, for reweaving this thesis and help in microarray printing and analysis.

Dr. Claus Hultschig and Dr. Ingo Fritz, Department of Molecular Genetics of the Max Planck Institute (MPI) in Berlin, for advices and practical lessons in fabrication of microarray biochips and PCR troubleshooting.

Dr. Robert Geffers, Group of Mucosal Immunity at the GBF, for being a member of my PhD committee, for support in production of microarray biochips and great help in analysis of microarray data.

Dr. Werner Tegge, CBIO at the GBF, for being a member of my PhD committee

Thanks to Ulrike Beutling (Dipl. Chem. Ibd), CBIO at the GBF, for synthesis of peptides on a novel polypropylene sheet support.

Susanne Daenicke (TA), Heidrun Lorenz-Küfner (TA), Edelgard Schmeißer (TA), CBIO at the GBF, for providing me with peptides synthesized on cellulose membranes and assisting me in caring out the chemical peptide SPOT synthesis.

Prof. Dr. Dr. Joachim Klose, Institute of Human Genetics, Charité University in Berlin, for providing me a list of proteins from mouse brains identified using 2D gel electrophoresis and mass-spectrometry technologies

Dr. David Lincoln for development of the Accessible Surface Scanner (ASS) software tool and discussion related to its improvement.

Dr. Helmut Blöcker, Department of Genome Analysis at the GBF, for allowance in using the high-throughput PCR facility in his laboratory.

Great thanks to Varsha Gupte (MSc), CBIO at the GBF, for being very supportive, great lab colleague, for very nice scientific and non-scientific “chats”, and great lab atmosphere.

Dr. Alexei Kurakin, the Buck Institute for Age Research in Novato (USA), for his assistance in learning Target Assisted Iterative Screening (TAIS) technology and great scientific discussion.

Great thanks to Dr. Cornelia Hunke, group of Conformational Protein Ligand Interactions (KPLI) at the GBF, for providing me with the GST-MenaEVH1 fusion protein, anti-GST and anti-EVH1antibodies.

In addition I would like thank to each member of the CBIO and KPLI research groups at the GBF, for having wonderful time during my studies in Braunschweig.

Thanks to my good friends, Jadwiga Jablonska and Andreas Koch, for having a great time together in Braunschweig and all my friends whom I have failed to mention here in person, for being supportive.

My very special thanks go to my wife Anna for her assistance in correcting and assembling of this thesis, her endless patience, understanding and support in all times and cases.

To my parents.

Moim Rodzicom
Mamo i Tato, dziękuję Wam za waszą milość i wsparcie w każdej chwili mojego życia.

CURRICULUM VITAE

Name:	Andrzej Swistowski
Date of birth:	05/15/1976
Place of birth:	Klodzko / Poland
Nationality:	Polish
Marital status:	married

Education:

1982-1990:	Primary school – Woliborz, Dolny Slask
1990-1994:	Secondary school - biology-chemistry profile Liceum Ogolnoksztalcace im.Henryka Sienkiewicza, Nowa Ruda, Dolny Slask
1995-1996:	Technical school of medical analytics – Wroclaw, Dolny Slask

Scientific education:

Sep 1996 – Jun 1999:	BA degree in biotechnology – The University of Wroclaw
Sep 1999 - July 2001:	MSc degree in biotechnology – The University of Wroclaw. The Institute of Immunology and Experimental Therapy of the Polish Academy of Science (PAN).
Feb 2001 - July 2001:	Socrates/Erasmus Fellowship. Graduate research project in the Department of Membrane Cell Biology Faculty of Medical Sciences at the University of Groningen (the Netherlands)
Oct 2001 - Apr 2005:	Ph.D. studies in the Department of Chemical Biology of the German Research Center for Biotechnology (GBF), Braunschweig (Germany)
Since Jul 2005:	Post Doctoral research fellowship at The Buck Institute for Age Research, Novato, CA, (USA)